



# Descripteurs couleur locaux invariants aux conditions d'acquisition

Xiaohu Song

## ► To cite this version:

Xiaohu Song. Descripteurs couleur locaux invariants aux conditions d'acquisition. Autre [cond-mat.other]. Université Jean Monnet - Saint-Etienne, 2011. Français. NNT : 2011STET4033 . tel-00719349

**HAL Id: tel-00719349**

**<https://theses.hal.science/tel-00719349>**

Submitted on 19 Jul 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE JEAN MONNET

N° attribué par la bibliothèque

# THÈSE

pour obtenir le grade de

**DOCTEUR de l'UNIVERSITÉ JEAN MONNET**

Spécialité : **Analyse et Traitement d'Images**

préparée au laboratoire **Hubert Curien**

dans le cadre de l'École Doctorale **SIS**

présentée et soutenue publiquement

par

**XIAOHU SONG**

le 08 Décembre 2011

Titre:

**Descripteurs couleur locaux invariants aux conditions  
d'acquisition**

Directeur de thèse: **M. Alain TRÉMEAU**

Co-directeur de thèse: **M. Damien MUSELET**

Jury

M. Ludovic MACAIRE,	Rapporteur
M. Azeddine BEGHDADI,	Rapporteur
Mme. Michèle GOUIFFES,	Examineur
M. Patrick LAMBERT,	Examineur
M. Dietrich PAULUS,	Examineur
M. Fabrice MERIAUDEAU,	Examineur

---

# RESUMÉ

La mise au point de descripteurs locaux discriminants est aujourd'hui une priorité dans de nombreuses applications comme la reconnaissance d'objets, le suivi d'objets, la reconstruction 3D ou l'estimation de mouvement. La problématique réside dans le fait que ces descripteurs doivent être invariants aux conditions d'acquisition tout en conservant un pouvoir discriminant important. Dans ce contexte, nous nous sommes intéressés à l'invariance des descripteurs locaux de la littérature. Nous les avons notamment catégorisé en fonction des hypothèses sur lesquelles repose leur invariance.

Ensuite, nous avons proposé des descripteurs locaux qui exploitent l'information de couleur dans les images. Nous avons montré que cette information peut être très pertinente lorsqu'elle est combinée à une information spatiale, à condition que son degré d'invariance soit contrôlé et adapté aux applications considérées. Ainsi, nous avons proposé un ensemble de descripteurs locaux couleur avec des degrés d'invariance différents.

Ainsi, nous introduisons tout d'abord deux nouveaux descripteurs qui caractérisent les distributions spatiales des couleurs dans les régions analysées. L'idée originale consiste à appliquer des transformations affines entre les coordonnées spatiales des pixels et leurs coordonnées couleur. En effet, chaque pixel étant ca-

---

ractérisé par 5 valeurs, 2 coordonnées spatiales  $xy$  dans l'image et 3 composantes couleur RVB, nous proposons de rechercher une transformation affine qui permet de transformer les coordonnées  $xy$  de tous les pixels de la région concernée en coordonnées RVB de ces pixels. Nous montrons que l'application de cette transformation aux coordonnées  $xy$  fournit des coordonnées dans l'espace RVB qui a un double avantage. D'une part, les coordonnées d'un seul pixel dépendent à la fois de toutes les couleurs présentes dans la région mais aussi de leur répartition spatiale. Quelques coordonnées permettent donc de résumer efficacement le contenu de la région. D'autre part, ces coordonnées présente une invariance totale à toute transformation affine appliquée dans l'espace image  $2D$  (invariance géométrique) et comme elles sont homogènes à des coordonnées couleur, nous pouvons leur procurer une invariance photométrique en leur appliquant des transformations affines particulières. Nous montrons que le degré d'invariance peut être contrôlé en fonction des besoins de l'application. Ces coordonnées nous permettent de définir le descripteur IVC (Image Vers Couleur). De manière similaire, nous évaluons une transformation affine de l'espace couleur à l'espace image et appliquons cette transformation aux coordonnées couleur. Les coordonnées obtenues par cette transformation sont invariantes à toute transformation affine appliquée dans l'espace couleur, elles présentent donc un degré d'invariance élevé aux variations photométriques. Ces coordonnées nous permettent de constituer le descripteur CVI (Couleur Vers Image). Nous montrons que ces deux descripteurs fournissent de très bons résultats dans le cadre de la reconnaissance d'objet et présentent une telle complémentarité que le descripteur obtenu par concaténation de IVC et CVI fournit de meilleurs résultats que la plupart des descripteurs couleur parus dans la littérature.

Ensuite, nous proposons un descripteur qui présente un degré d'invariance plus élevé que les deux précédents puisqu'il n'est pas sensible aux transformations non-linéaires des couleurs modélisées par des fonctions croissantes appliquées indépendamment sur chaque composante couleur. Pour cela, nous exploitons les

---

mesures de rang des pixels dans les images. Ces mesures de rang sont les positions normalisées entre 0 et 1 des pixels dans les images lorsqu'ils sont classés dans l'ordre croissant de leur niveau de composante. Chaque pixel est caractérisé par 3 mesures de rang, chacune obtenue par le classement selon une composante couleur R, V ou B. Ces mesures de rang ont déjà été utilisées dans plusieurs travaux mais l'originalité de notre approche réside dans le fait d'exploiter le coefficient de Kendall qui est une mesure mathématique qui caractérise la corrélation entre les mesures de rang de deux variables. Nous avons adapté ce coefficient à notre problématique de manière à extraire les corrélations entre les mesures de rang de différents ensembles de pixels dans la région à caractériser.

De plus, nous utilisons les corrélations entre mesures de rang obtenues pour différentes composantes couleur. Ceci nous a permis de proposer un descripteur lui aussi très compact qui présente un degré d'invariance photométrique assez élevé.

Enfin, nous abordons le problème de la caractérisation locale d'images par auto-similarités. L'objectif de ces approches est de pouvoir reconnaître des objets dont les caractéristiques locales peuvent être très différentes mais qui partagent cependant des spécificités globales. Dans ce cas, l'idée consiste à comparer les régions locales d'une même image entre elles et à caractériser le contenu de l'image par la répartition des mesures de similarités obtenues. Comme les mesures de similarité sont calculées entre régions d'une même image, nous parlons d'auto-similarités. Cette approche permet de caractériser de manière similaire les images aux contenus très divers mais elle peut aussi être adaptée à la catégorisation d'objets pour laquelle les objets d'une même classe (voiture, vélo, fleurs, ...) peuvent présenter des couleurs ou des textures différentes. Par exemple, un vélo bleu et un vélo vert appartiennent tous les deux à la classe des vélos. Des descripteurs couleur ne peuvent pas permettre de classer ces images dans la même classe, mais la distribution spatiale de leurs auto-similarités peut aider. Enfin, cette approche peut aussi être utilisée lorsque les variations des couleurs entre deux images d'un

---

même objet ne peuvent pas être modélisées par une fonction mathématique. Dans ce cadre, nous proposons donc un descripteur invariant à la rotation qui permet de retrouver très vite les régions similaires au sein d'une même image.

Nous testons nos descripteurs locaux dans le cadre de la comparaison de régions locales extraites d'images acquises sous des conditions très variées ainsi que dans des applications de catégorisation d'objets. Nous comparons les performances de nos descripteurs avec celles fournies par les descripteurs locaux couleur qui fournissent les meilleurs résultats dans la littérature. Nous verrons que, malgré leur compacité, nos descripteurs présentent des pouvoirs discriminants supérieurs à ceux des autres descripteurs testés sur la majorité des bases considérées.

**Mots clés : descripteurs locaux, reconnaissance d'objets, invariants, transformations affines, pouvoir discriminant, IVC, CVI, mesures de rang, mesures de similarités.**

---

# REMERCIEMENTS

En premier lieu, je tiens à remercier toutes les personnes de l'équipe de "color content aware processing" du laboratoire Hubert Curien de l'université Jean Monnet, spécialement les professeurs qui m'ont formé et encadré au cours de ces trois années de thèse.

Je tiens à exprimer ma gratitude et mes profonds remerciements à Mr. Damien Muselet, directeur de thèse, pour la qualité de son encadrement. Sa grande disponibilité, sa compétence et ses nombreux conseils prodigués ont été le principal moteur de l'avancement de la thèse, dans ses grandes lignes comme dans la résolution des différents problèmes scientifiques rencontrés. J'ai aussi particulièrement apprécié les relations chaleureuses qu'il a su insuffler au sein de l'équipe de Alain Trémeau.

Je tiens aussi particulièrement à exprimer mes sincères à mon directeur de thèse du professeur Alain Trémeau qui m'a accepté et m'a aidé durant mes travaux. Qu'ils soient assurés de ma reconnaissance pour son implication, son soutien scientifique et humain et sa relecture efficace de l'écrit de la thèse. Il a toujours été présent, disponible et m'a donné des conseils durant toute la thèse. Je lui remercie pour les discussions intéressantes que nous avons eues et pour son amitié.

---

Messieurs Ludovic Macaire et Azeddine Beghdadi m'ont fait l'honneur d'accepter de rapporter cette thèse, et de se déplacer de loin pour participer à la soutenance. Je leur en exprime ma profonde gratitude. Mes remerciements vont également aux autres personnes du jury, Madame Michèle GOUIFFES, Mr Patrick LAMBERT, Mr Dietrich PAULUS et Mr Fabrice MERIAUDEAU.

J'adresse mes remerciements aux autres personnes du DISCo (Dipartimento di Informatica, Sistemistica e Comunicazione) en Italie qui j'ai eu le plaisir de collaborer : Mr. Schettini raymondo et Mr. Claudio Cusano qui m'ont donnés l'occasion de mettre incontestablement en valeur des outils présentés au cours de la collaboration.

J'adresse finalement mes remerciements les plus affectueux à tous les membres de ma famille. Aucun mot n'est assez fort pour exprimer mes sentiments pour eux.

Et un grand MERCI à tous ceux qui m'ont aidé pendant ma préparation de la thèse.



---

# TABLE DES MATIÈRES

<b>RESUMÉ</b>	<b>1</b>
<b>REMERCIEMENTS</b>	<b>5</b>
1 Introduction générale . . . . .	11
2 Les descripteurs couleur invariants de la littérature . . . . .	20
2.1 Introduction . . . . .	22
2.2 Les hypothèses de base [80] . . . . .	23
2.2.1 Hypothèses sur la formation de la couleur . . . . .	23
2.2.2 Hypothèse sur les propriétés de réflexion des éléments de surface . . . . .	27
2.2.3 Hypothèses sur les sensibilités des capteurs de la caméra .	27
2.2.4 Hypothèses sur les caractéristiques de l'illuminant . . . .	29
2.2.5 Hypothèses sur le modèle de variations photométriques et radiométriques . . . . .	31
2.3 Les caractéristiques couleur invariantes . . . . .	35
2.3.1 Les rapports inter et intra composantes couleur . . . . .	36

2.3.2	Les transformations fondées sur une analyse des distributions colorimétriques . . . . .	47
2.3.3	Les dérivées invariantes . . . . .	50
2.3.4	Synthèse sur les hypothèses et les invariances des caractéristiques couleur . . . . .	62
2.4	Les descripteurs locaux couleur . . . . .	68
2.4.1	Concaténation de deux types de descripteurs . . . . .	69
2.4.2	Deux étapes successives pour la comparaison d'images . . . . .	71
2.4.3	Comparaisons parallèles . . . . .	74
2.4.4	Descripteurs spatio-colorimétriques . . . . .	75
2.5	Conclusion . . . . .	78
3	Invariance par transformations affines entre espaces couleur et image . . . . .	81
3.1	Introduction . . . . .	83
3.2	Transformations affines entre deux espaces différents . . . . .	84
3.2.1	Solution des moindres carrés . . . . .	85
3.2.2	Pouvoir discriminant . . . . .	90
3.2.3	Invariance et covariance . . . . .	92
3.2.4	Contraintes sur la transformation $T_{OVD}$ . . . . .	94
3.3	De l'espace image vers l'espace couleur . . . . .	99
3.3.1	Illustration . . . . .	99
3.3.2	Le descripteur IVC . . . . .	101
3.3.3	Les contraintes géométriques et couleur d'IVC . . . . .	104
3.4	De l'espace couleur à l'espace image . . . . .	108
3.4.1	Illustration . . . . .	109
3.4.2	Le descripteur CVI . . . . .	109
3.5	Conclusion . . . . .	113
4	L'invariance par mesures de rangs . . . . .	115
4.1	Introduction . . . . .	116
4.2	Mesures de rangs . . . . .	117

## TABLE DES MATIÈRES

4.3	Coefficients de corrélation de rangs de Kendall . . . . .	118
4.4	Corrélation de rangs entre sous-ensembles . . . . .	121
4.5	Le descripteur local fondé sur les corrélations de rang couleur . .	126
4.6	Conclusion . . . . .	127
5	L'invariance par auto-similarités . . . . .	129
5.1	Introduction . . . . .	130
5.2	Les descripteurs fondés sur l'auto-similarité . . . . .	133
5.2.1	Descripteur local d'auto-similarités [95] . . . . .	133
5.2.2	Descripteur global d'auto-similarités [24] . . . . .	136
5.3	Calcul d'auto-similarités invariantes à la rotation . . . . .	140
5.4	Représentation locale de la distribution spatiale des auto-similarités	145
5.5	Conclusion . . . . .	146
6	Résultats expérimentaux . . . . .	148
6.1	Introduction . . . . .	149
6.2	Les bases d'images . . . . .	149
6.2.1	La base d'images pour l'appariement . . . . .	149
6.2.2	Les bases d'images pour la classification . . . . .	151
6.3	Les descripteurs testés . . . . .	159
6.3.1	Propriétés théoriques d'invariance . . . . .	159
6.3.2	Compacités . . . . .	160
6.4	Appariement de régions locales . . . . .	161
6.4.1	Paramètres de IVC et CVI . . . . .	162
6.4.2	Comparaison avec les descripteurs de la littérature . . . .	169
6.5	Classification d'objets . . . . .	181
6.6	Conclusion . . . . .	186
7	Conclusion générale et perspectives . . . . .	187
	Bibliographie . . . . .	193
	Annexes . . . . .	213
	Annexe A Sac de mots . . . . .	214

## TABLE DES MATIÈRES

---

Annexe B	SIFT : Scale-invariant feature transform . . . . .	217
Annexe C	La liste des publications en rapport avec la thèse . . . . .	220

---

---

# CHAPITRE 1

---

## INTRODUCTION GÉNÉRALE

*"A Picture of many colors proclaims images of many Thoughts", Donna A.*

Lorsque l'on souhaite reconnaître et localiser un objet particulier dans une image [17], recalcr deux images acquises sous différents points de vue [108], calibrer une caméra [13], reconstruire en 3 dimensions une scène acquise sous différents points de vue [92], estimer le mouvement dans une séquence [12], corriger les couleurs entre 2 images de la même scène [59], suivre un objet dans une vidéo [110] ou même classer des objets ou des images dans différentes catégories (vélo, personne, chien, ...) [98], une étape pratiquement incontournable est celle qui consiste à extraire des descripteurs de petites régions locales. En effet, la description locale est devenue aujourd'hui un point clé de nombreuses applications de vision et nous en avons fait l'objet de cette thèse de doctorat.

Une étape préalable à la description de régions locales est l'extraction de ces régions dans les images. Il existe de nombreux détecteurs de points d'intérêt et de régions d'intérêt [67, 74, 97] dont le but est d'extraire des points ou des régions qui ont des propriétés particulières dans les images (coins, contours, régions

saillantes, régions homogènes...) et qui sont stables en cas de variations des conditions d'acquisition. Au cours de nos travaux, nous nous sommes concentrés sur la description des régions fournies par ces détecteurs sans nous attarder sur les détecteurs eux-mêmes. Nous partons donc du principe que les régions ont été détectées dans les images et que notre tâche est de caractériser leur contenu.

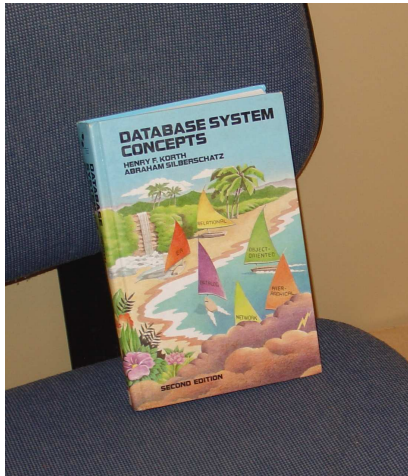
La plupart des capteurs d'acquisition (appareils photo, caméras, téléphones, scanners, ...) fournissant des images couleur, nous nous sommes focalisés sur la prise en compte de l'information colorimétrique dans les descripteurs locaux. Nous verrons que cette information peut être très discriminante même si elle reste très sensible aux variations des conditions d'acquisition. En effet, dans la plupart des applications pré-citées, les conditions d'acquisition ne sont ni contrôlées, ni connues. Ainsi, lorsque nous recherchons un objet dans une base d'image par exemple, nous avons une photo de cet objet (image gauche dans la Figure 1) que nous comparons à toutes les images de la base considérée et l'image de la base qui contient cet objet (image droite dans la Figure 1) a certainement été acquise sous des conditions différentes. C'est à dire que dans cette application, comme dans la plupart des autres, entre les deux prises de vue d'une même scène ou d'un même objet, nous considérons que les variations suivantes peuvent avoir lieu :

- variation du point de vue,
- différentes positions 3D de l'objet (facteur d'échelle, rotation 3D, translation) dans la scène,
- occultation d'une partie de l'objet,
- systèmes d'acquisition différents,
- systèmes d'éclairages (couleur, direction, nombre) différents,
- apparition d'ombres ou de reflets.

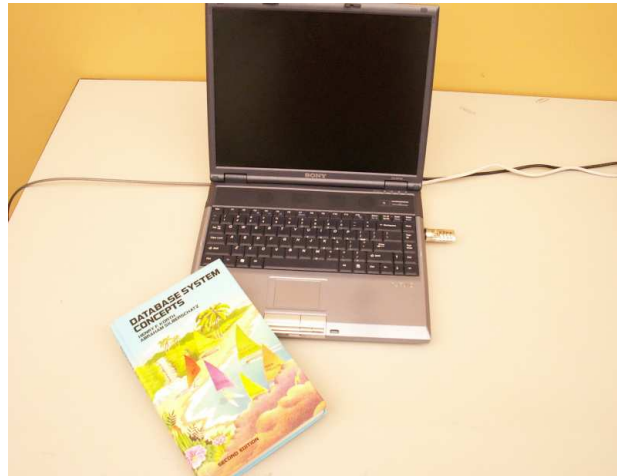
Dans ce cas, une étape de normalisation photométrique et géométrique doit être appliquée afin de rendre les descripteurs moins sensibles aux variations éventuelles des conditions d'acquisition.

L'apport de la couleur dans les descripteurs locaux ainsi que leur invariance aux conditions d'acquisition sont donc les deux thèmes majeurs traités dans cette thèse. La principale contribution de nos travaux réside dans la mise au point de descripteurs locaux compacts, qui caractérisent la distribution spatiale des couleurs localement et qui présentent une invariance à une large plage de variations des conditions d'acquisition.

De sorte à évaluer la pertinence de ces descripteurs par rapport à ceux de la littérature, il convient de connaître les critères de qualité d'un descripteur. Nous proposons d'en retenir 3 principaux. Le premier est son pouvoir discriminant, c'est à dire la capacité du descripteur à différencier le cas où deux régions représentent les mêmes éléments de surface d'une scène du cas où elles représentent des éléments de surface différents. Le second est son degré d'invariance qui dépend de sa sensibilité aux variations des conditions d'acquisition. Le troisième est sa compacité qui est liée à la place mémoire occupée par celui-ci. Nous avons déjà noté que ces critères ne sont pas complètement indépendants. En effet, nous verrons dans ce manuscrit, par exemple, qu'il existe un lien étroit entre le degré d'invariance et le pouvoir discriminant d'un descripteur. Notamment, nous montrerons que le degré d'invariance donné à un descripteur local ne doit pas être supérieur au degré d'invariance requis par l'application, car tout degré d'invariance superflu risque de diminuer le pouvoir discriminant de ce descripteur. Par exemple, lorsque toutes les images d'une base sont acquises par le même système d'acquisition, il n'est pas nécessaire d'utiliser un descripteur invariant aux capteurs d'acquisition. Toute la difficulté dans la mise au point d'un descripteur est de savoir gérer ce compromis entre pouvoir discriminant et degré d'invariance. Pour cela, il est indispensable de connaître l'impact qu'aura la variation d'un paramètre d'acquisition (capteurs, couleur de l'éclairage, ...) sur les couleurs des images afin de doser



Une image de l'objet recherché



Une image de la base contenant cet objet

FIGURE 1.1 – Conditions d'acquisition non contrôlées dans le cadre de la recherche d'objets.

au mieux le degré d'invariance du descripteur choisi. L'objectif de cette thèse a donc été double. D'une part, analyser la formation de la couleur dans les images pour connaître le rôle de chaque élément entrant en jeu et ainsi prédire l'impact de variations potentielles des conditions d'acquisition sur les couleurs. D'autre part, proposer un ensemble de descripteurs locaux couleur dont les degrés d'invariance sont différents et connus afin d'avoir un descripteur adapté à chaque situation.

Dans ce manuscrit, nous allons dans un premier temps revenir sur la plupart des descripteurs locaux qui prennent en compte la couleur. Après une analyse de la littérature, il apparaît que la plupart des descripteurs locaux (SIFT, SURF, GLOH, ...) sont fondés sur les gradients locaux et exploitent donc principalement l'information de forme. Or, les gradients locaux présentent un certains nombre d'inconvénients. D'une part, ils sont sensibles aux bruits dans les images ainsi qu'aux fortes variations de point de vue. Ensuite, ils constituent une information peu compacte et peuvent difficilement être résumés par quelques valeurs. Enfin, les descripteurs qui exploitent les gradients ont été initialement définis pour caractériser des régions en niveaux de gris et leur extension à la couleur consiste sim-



plement à extraire ces gradients sur chaque composante couleur, négligeant ainsi l'aspect vectoriel de la couleur. Dans ce manuscrit, nous soutenons l'idée que les descripteurs fondés sur le gradient ne constituent pas la meilleure structure pour incorporer une information couleur et nous proposons plutôt de caractériser les régions locales par la distribution spatiale des couleurs qui les composent.

Nous allons analyser de quelle manière la couleur peut être normalisée avant d'être exploitée par ces descripteurs. Cette normalisation passe par la modélisation des variations des couleurs en cas de changements des conditions d'acquisition par une fonction mathématique. Cette fonction n'existe pas en pratique et si l'on souhaite l'approcher, il est nécessaire de poser un certain nombre d'hypothèses sur les différents paramètres qui entrent en jeu dans la formation de la couleur. Dans le chapitre 2 de ce manuscrit, nous dresserons une liste de la plupart des hypothèses utilisées dans la littérature. Il est clair que chaque hypothèse posée va, d'un côté, simplifier la forme de la fonction recherchée et d'un autre côté restreindre le champs d'application des composantes couleur invariantes déduites de cette fonction. Autrement dit, plus les hypothèses seront restrictives, plus la fonction à estimer sera simple et moins le degré d'invariance du descripteur sera élevé. Ainsi, pour les hypothèses les plus restrictives, nous verrons que la transformation des couleurs en cas de changement des conditions d'acquisition peut être assimilée à une transformation affine dans l'espace  $3D$  des couleurs et que le nombre de degrés de liberté de cette transformation (matrice diagonale, translation, matrice  $3 \times 3$ , ...) est d'autant plus faible que les hypothèses seront restrictives. Si les hypothèses sont moins restrictives, les variations des conditions d'acquisition sont trop complexes pour être assimilées à une transformation affine et une alternative consiste plutôt à les modéliser par des fonctions croissantes appliquées sur chacune des composantes. Cette modélisation est plus souple puisqu'elle peut être non linéaire. Enfin, il se peut que les variations ne puissent pas être modélisées par une fonction.

Chacun des chapitres 3, 4 et 5 va se placer dans ces conditions particulières.

En effet, dans le chapitre 3, nous allons proposer deux descripteurs locaux couleur complémentaires qui présentent des invariances aux transformations affines appliquées dans l'espace couleur. Dans le chapitre 4, le descripteur proposé est fondé sur les mesures de rang des pixels dans les images qui sont invariants à l'application de fonctions croissantes sur les composantes couleur. Enfin, le chapitre 5 aborde les descripteurs qui exploitent les auto-similarités dans les images et qui ne nécessitent pas une modélisation des transformations en cas de variations des conditions d'acquisition.

Ainsi, le chapitre 3, qui est le cœur de cette thèse, introduit deux nouveaux descripteurs qui caractérisent les distributions spatiales des couleurs dans les régions analysées. L'idée originale consiste à appliquer des transformations affines entre les coordonnées spatiales des pixels et leurs coordonnées couleur. En effet, chaque pixel étant caractérisé par 5 valeurs, 2 coordonnées spatiales  $xy$  dans l'image et 3 composantes couleur  $RVB$ , nous proposons de rechercher une transformation affine qui permet de transformer les coordonnées  $xy$  de tous les pixels de la région concernée en coordonnées  $RVB$  de ces pixels. Nous montrons que l'application de cette transformation aux coordonnées  $xy$  fournit des coordonnées dans l'espace  $RVB$  qui ont un double avantage. D'une part, les coordonnées d'un seul pixel dépendent à la fois de toutes les couleurs présentes dans la région mais aussi de leur répartition spatiale. Quelques coordonnées permettent donc de résumer efficacement le contenu de la région. D'autre part, ces coordonnées présentent une invariance totale à toute transformation affine appliquée dans l'espace image  $2D$  (invariance géométrique) et comme elles sont homogènes à des coordonnées couleur, nous pouvons leur procurer une invariance photométrique en leur appliquant des transformations affines particulières. Nous montrons que le degré d'invariance peut être contrôlé en fonction des besoins de l'application. Ces coordonnées nous permettent de définir le descripteur IVC (Image Vers Couleur). De manière similaire, nous évaluons une transformation affine de l'espace couleur à l'espace image et appliquons cette transformation aux coordonnées couleur. Les coordon-

nées obtenues par cette transformation sont invariantes à toute transformation affine appliquée dans l'espace couleur, elles présentent donc un degré d'invariance élevé aux variations photométriques. Ces coordonnées nous permettent de constituer le descripteur CVI (Couleur Vers Image). Nous montrons dans le chapitre 6 que les deux descripteurs proposés dans ce chapitre fournissent de très bons résultats dans le cadre de l'appariement de régions et présentent une telle complémentarité que le descripteur obtenu par concaténation de IVC et CVI fournit de meilleurs résultats que la plupart des descripteurs couleur parus dans la littérature.

Dans le chapitre 4, nous proposons un descripteur qui présente un degré d'invariance plus élevé que les deux précédents puisqu'il n'est pas sensible aux transformations non-linéaires des couleurs modélisées par des fonctions croissantes appliquées indépendamment sur chaque composante couleur. Pour cela, nous exploitons les mesures de rang des pixels dans les images. Ces mesures de rang sont les positions normalisées entre 0 et 1 des pixels dans les images lorsqu'ils sont classés dans l'ordre croissant de leur niveau de composante. Chaque pixel est caractérisé par 3 mesures de rang, chacune obtenue par le classement selon une composante couleur  $R$ ,  $V$  ou  $B$ . Ces mesures de rang ont déjà été utilisées dans plusieurs travaux [28, 82, 83] mais l'originalité de notre approche réside dans le fait d'exploiter le coefficient de Kendall qui est une mesure mathématique qui caractérise la corrélation entre les mesures de rang de deux variables. Nous avons adapté ce coefficient à notre problématique de manière à extraire les corrélations entre les mesures de rang de différents ensembles de pixels dans la région à caractériser. De plus, nous utilisons les corrélations entre mesures de rang obtenues pour différentes composantes couleur. Ceci nous a permis de proposer un descripteur lui aussi très compact qui présente un degré d'invariance photométrique assez élevé.

Dans le chapitre 5, nous avons abordé le problème de la caractérisation locale d'images par auto-similarités [19, 24, 95]. L'objectif de ces approches est de pouvoir reconnaître des objets dont les caractéristiques locales peuvent être très

différentes mais qui partagent cependant des spécificités globales. Les images de la figure 1.2 illustrent cette problématique. Dans ce cas, l'idée consiste à comparer les régions locales d'une même image entre elles et à caractériser le contenu de l'image par la répartition des mesures de similarités obtenues. Comme les mesures de similarité sont calculées entre régions d'une même image, nous parlons d'auto-similarités. Cette approche permet de caractériser de manière similaire les images de la figure 1.2 mais elle peut aussi être adaptée à la catégorisation d'objets pour laquelle les objets d'une même classe (voiture, vélo, fleurs, ...) peuvent présenter des couleurs ou des textures différentes. Par exemple, un vélo bleu et un vélo vert appartiennent tous les deux à la classe des vélos. Des descripteurs couleur ne peuvent pas permettre de classer ces images dans la même classe, mais la distribution spatiale de leurs auto-similarités peut aider. Enfin, cette approche peut aussi être utilisée lorsque les variations des couleurs entre deux images d'un même objet ne peuvent pas être modélisées par une fonction mathématique.

Dans ce contexte, les approches classiques [95] comparent les régions locales en évaluant les corrélations entre fenêtres pixel à pixel. Ceci est très coûteux en temps de calcul et n'est pas invariant à la rotation entre les régions. Or, dans une image, nous montrons qu'il peut être très intéressant de retrouver des auto-similarités entre régions même si celles-ci ne sont pas orientées de la même manière. Dans ce cadre, nous proposons donc un descripteur invariant à la rotation qui permet de retrouver très vite les régions similaires au sein d'une même image. Ce descripteur est une extension du descripteur IVC introduit dans le chapitre 3 de ce manuscrit.

Dans le chapitre 6, nous exposerons les résultats expérimentaux. Nous allons tester nos descripteurs locaux dans le cadre de la comparaison de régions locales extraites d'images acquises sous des conditions très variées ainsi que dans des applications de catégorisation d'objets. Nous comparerons les performances de nos descripteurs avec celles fournies par les descripteurs locaux couleur qui fournissent les meilleurs résultats dans la littérature. Nous verrons que, malgré leur



FIGURE 1.2 – Ces images ont des caractéristiques locales différentes mais des répartitions d’auto-similarités similaires.

compacité, nos descripteurs présentent des pouvoirs discriminants supérieurs à ceux des autres descripteurs testés sur la majorité des bases considérées.

Nous terminerons ce mémoire en concluant par rapport aux travaux qui ont été réalisés et nous proposerons les pistes futures qui nous semblent les plus appropriées.



---

---

## CHAPITRE 2

---

# LES DESCRIPTEURS COULEUR INVARIANTS DE LA LITTÉRATURE

### Table des matières

---

<b>2.1</b>	<b>Introduction . . . . .</b>	<b>22</b>
<b>2.2</b>	<b>Les hypothèses de base [80] . . . . .</b>	<b>23</b>
2.2.1	Hypothèses sur la formation de la couleur . . . . .	23
2.2.2	Hypothèse sur les propriétés de réflexion des éléments de surface . . . . .	27
2.2.3	Hypothèses sur les sensibilités des capteurs de la caméra	27
2.2.4	Hypothèses sur les caractéristiques de l'illuminant . .	29
2.2.5	Hypothèses sur le modèle de variations photométriques et radiométriques . . . . .	31
<b>2.3</b>	<b>Les caractéristiques couleur invariantes . . . . .</b>	<b>35</b>

## CHAPITRE 2. LES DESCRIPTEURS COULEUR INVARIANTS DE LA LITTÉRATURE

---

2.3.1	Les rapports inter et intra composantes couleur . . . .	36
2.3.2	Les transformations fondées sur une analyse des distributions colorimétriques . . . . .	47
2.3.3	Les dérivées invariantes . . . . .	50
2.3.4	Synthèse sur les hypothèses et les invariances des caractéristiques couleur . . . . .	62
<b>2.4</b>	<b>Les descripteurs locaux couleur . . . . .</b>	<b>68</b>
2.4.1	Concaténation de deux types de descripteurs . . . . .	69
2.4.2	Deux étapes successives pour la comparaison d'images	71
2.4.3	Comparaisons parallèles . . . . .	74
2.4.4	Descripteurs spatio-colorimétriques . . . . .	75
<b>2.5</b>	<b>Conclusion . . . . .</b>	<b>78</b>

---



## 2.1 Introduction

Dans ce chapitre, nous proposons de dresser un état de l'art des descripteurs couleur locaux. Comme nous travaillons sous des conditions d'acquisition non contrôlées, nous allons notamment nous intéresser à la part d'invariance que présentent ces descripteurs. Nous verrons que dans la plupart des cas, les descripteurs locaux invariants sont obtenus grâce à deux étapes successives. La première est un prétraitement des images couleur qui consiste à transformer les composantes RVB des pixels en composantes invariantes. La seconde extrait de ces images invariantes des descripteurs locaux.

La notion de "caractéristiques couleurs invariantes" repose sur des modèles de variations photométriques et radiométriques qui associent à ces variations une fonction mathématique. Pour expliquer ce qu'est un modèle de variations photométriques et radiométriques, plaçons nous dans le cadre de la reconnaissance d'objet dans des bases d'images et considérons deux images couleur requête  $\mathbf{I}_{req}$  et candidate  $\mathbf{I}_{cand}$  qui représentent le même objet. Soient  $P_{req}$  et  $P_{cand}$ , deux pixels appartenant respectivement aux images  $\mathbf{I}_{req}$  et  $\mathbf{I}_{cand}$ , tels que le même élément de surface de l'objet représenté par les deux images se projette sur ces deux pixels. Supposons que cette surface élémentaire soit éclairée et observée différemment lors des acquisitions des images requête et candidate. Alors, le pixel  $P_{req}$  est caractérisé par la couleur  $\mathbf{C}(P_{req})$  dans l'image  $\mathbf{I}_{req}$  et le pixel  $P_{cand}$  est caractérisé par la couleur  $\mathbf{C}(P_{cand})$  dans l'image  $\mathbf{I}_{cand}$ . Un modèle de variations photométriques et radiométriques a pour but de définir la transformation  $\mathbf{F}_{ill}$  entre les couleurs associées à ces deux pixels :

$$\mathbf{C}(P_{cand}) = \mathbf{F}_{ill}(\mathbf{C}(P_{req})). \quad (2.1)$$

Il faut noter qu'une telle fonction n'existe pas, ne serait-ce qu'à cause du phénomène de métamérisme qui explique que deux surfaces de couleurs identiques

sous un éclairage donné peuvent être caractérisées par deux couleurs différentes sous un autre éclairage. Cependant, les modèles de variations photométriques et radiométriques tentent d’approcher cette relation par une fonction en posant un certain nombre d’hypothèses. Nous proposons de présenter dans la seconde partie de ce chapitre la plupart des hypothèses utilisées pour mettre au point des modèles de variations photométriques et radiométriques. Dans la troisième partie, nous dresserons un état de l’art des principales caractéristiques couleur invariantes qui reposent sur ces hypothèses. La quatrième partie présentera les différentes manières d’introduire une information couleur, invariante ou non, dans les descripteurs couleur locaux.

## **2.2 Les hypothèses de base [80]**

Les modèles de variations photométriques et radiométriques sont fondés sur différentes hypothèses qui portent sur la formation de la couleur, sur les propriétés de réflexion des éléments de surface, sur les sensibilités des capteurs de la caméra, sur l’illuminant qui éclaire la scène ou sur le modèle de variations photométriques et radiométriques, en lui-même. Nous nous proposons de lister dans les sous-sections suivantes l’ensemble des hypothèses qui conduisent aux modèles de variations photométriques et radiométriques les plus classiques.

### **2.2.1 Hypothèses sur la formation de la couleur**

La plupart des modèles de formation de la couleur repose sur l’hypothèse qu’un stimulus couleur qui atteint un matériau est réfléchi en partie immédiatement par la surface (réflexion de surface ou spéculaire) et que l’autre partie du stimulus pénètre le matériau, est réfléchi de manière aléatoire par les pigments qui constituent le matériau et est remis par le matériau suivant une direction, elle aussi, aléatoire (réflexion diffuse, cf. Figure 2.1). Ceci explique que la réflexion spéculaire ne soit pas perçue de manière identique selon la position de l’observa-

teur ou du capteur. Dans les deux cas, le spectre du stimulus original est modifié avant d'être transmis.

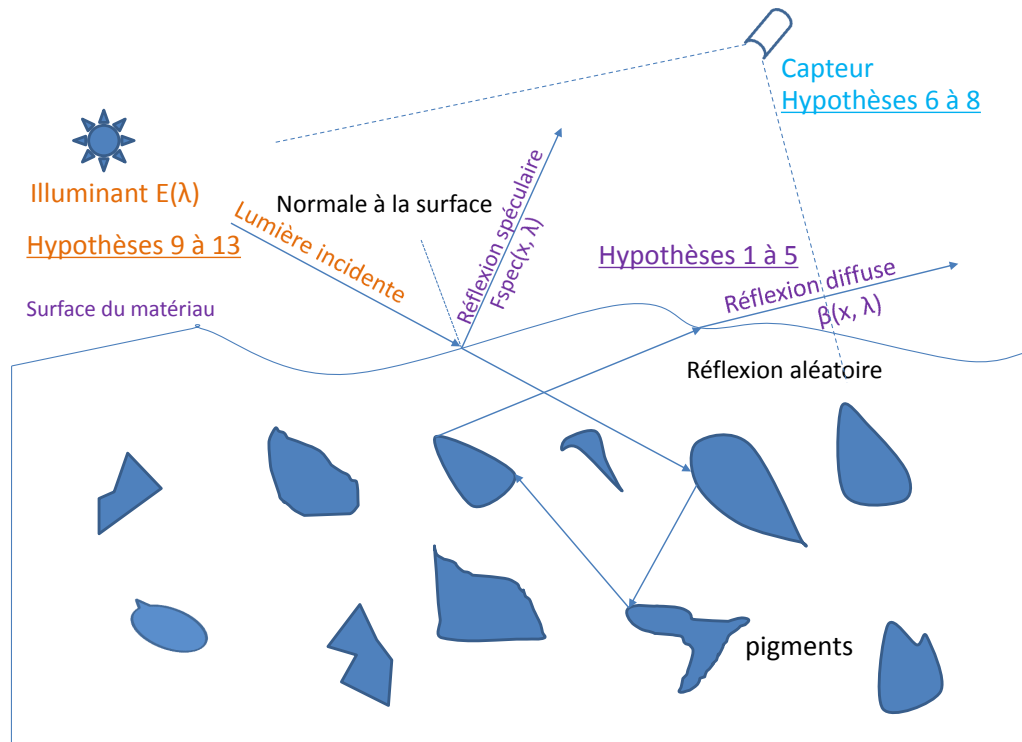


FIGURE 2.1 – Réflexion diffuse et spéculaire d'un rayon lumineux. Chaque élément est assujéti à diverse hypothèses détaillées dans ce chapitre.

Le stimulus couleur réfléchi est alors transmis aux capteurs d'acquisition dont les sensibilités spectrales sont notées  $k(\lambda)$ ,  $k = \{R, V, B\}$ . Si nous notons  $C(\mathbf{x}, \lambda)$  le stimulus couleur réfléchi par une surface élémentaire, nous rappelons que nous obtenons les composantes couleur du pixel  $P$  qui "observe" cette surface élémentaire en intégrant le produit des sensibilités de chaque capteur et du spectre du

stimulus sur l'intervalle de sensibilité de chaque capteur :

$$\begin{cases} C^R(P) = \int_{\lambda} R(\lambda)C(\mathbf{x}, \lambda)d\lambda, \\ C^V(P) = \int_{\lambda} V(\lambda)C(\mathbf{x}, \lambda)d\lambda, \\ C^B(P) = \int_{\lambda} B(\lambda)C(\mathbf{x}, \lambda)d\lambda. \end{cases} \quad (2.2)$$

Les trois modèles de formation de la couleur les plus couramment utilisés sont le modèle de Kubelka-Munk, le modèle de Shafer et le modèle Lambertien.

### **Hypothèse 1 : Modèle de Kubelka-Munk de formation de la couleur**

Nous considérons un matériau dont la réflectance diffuse à la position (dans l'espace)  $\mathbf{x}$  est  $\beta(\mathbf{x}, \lambda)$  et dont la réflectance spéculaire (ou de Fresnel) à la même position est  $F_{spec}(\mathbf{x}, \lambda)$ .

Le modèle de Kubelka-Munk [63] suppose que le stimulus couleur  $C_{Kub-Munk}(\mathbf{x}, \lambda)$  réfléchi par cette surface élémentaire, éclairée par un illuminant de répartition spectrale relative d'énergie  $E(\mathbf{x}, \lambda)$  est exprimé par :

$$C_{Kub-Munk}(\mathbf{x}, \lambda) = (1 - F_{spec}(\mathbf{x}, \lambda))^2 \cdot \beta(\mathbf{x}, \lambda) \cdot E(\mathbf{x}, \lambda) + F_{spec}(\mathbf{x}, \lambda) \cdot E(\mathbf{x}, \lambda). \quad (2.3)$$

### **Hypothèse 2 : Modèle de Shafer de formation de la couleur**

D'après le modèle de Shafer [94] ou modèle dichromatique, le stimulus de couleur  $C_{Shafer}(\mathbf{x}, \lambda)$  réfléchi par cette même surface élémentaire, éclairée par un illuminant de répartition spectrale relative d'énergie  $E(\mathbf{x}, \lambda)$  peut être exprimé par :

$$C_{Shafer}(\mathbf{x}, \lambda) = m_{diff}(\theta) \cdot \beta(\mathbf{x}, \lambda) \cdot E(\mathbf{x}, \lambda) + m_{spec}(\theta, \alpha) \cdot F_{spec}(\mathbf{x}, \lambda) \cdot E(\mathbf{x}, \lambda), \quad (2.4)$$

où  $m_{diff}$  et  $m_{spec}$  représentent la dépendance par rapport aux directions d'éclairage  $\theta$  et d'observation  $\alpha$ .

Les modèles de Shafer et de Kubelka-Munk sont similaires dans le sens où ils décomposent la réflexion en une réflexion spéculaire et une réflexion diffuse. Une telle décomposition d'un stimulus de couleur en deux termes correspondant à deux phénomènes physiques différents a été également vérifiée par Beckmann [14].

Cependant, ces deux modèles supposent que la lumière qui est "projetée" sur un pixel n'est émise que par une source ponctuelle. Dans l'article [94], Shafer propose d'étendre son modèle dichromatique en ajoutant un terme de lumière ambiante  $L_a(\lambda)$  qui est le même pour toutes les surfaces d'une scène observée et ne dépend donc pas de la position  $\mathbf{x}$ . Ce terme permet de prendre en compte plusieurs phénomènes comme l'influence des réflexions des surfaces élémentaires autour de la surface observée, la sensibilité des capteurs aux ondes infra-rouge ou la diffraction de la lentille [94]. Le stimulus couleur réfléchi complet devient alors :

$$C_{extended-Shafer}(\mathbf{x}, \lambda) = m_{diff}(\theta) \cdot \beta(\mathbf{x}, \lambda) \cdot E(\mathbf{x}, \lambda) + m_{spec}(\theta, \alpha) \cdot F_{spec}(\mathbf{x}, \lambda) \cdot E(\mathbf{x}, \lambda) + L_a(\lambda). \quad (2.5)$$

### Hypothèse 3 : Modèle Lambertien de formation de la couleur

Lorsque la surface observée est matte, la réflectance spéculaire peut être négligée. Dans ce cas précis, le modèle Lambertien estime le stimulus de couleur  $C_{Lambert}(\mathbf{x}, \lambda)$  réfléchi par la surface élémentaire considérée, éclairée par une source lumineuse de distribution spectrale  $E(\mathbf{x}, \lambda)$  de la manière suivante :

$$C_{Lambert}(\mathbf{x}, \lambda) = \beta(\mathbf{x}, \lambda) \cdot E(\mathbf{x}, \lambda). \quad (2.6)$$

### 2.2.2 Hypothèse sur les propriétés de réflexion des éléments de surface

#### Hypothèse 4 : Réflexion spéculaire neutre

L'analyse du spectre des réflectances spéculaires de nombreux matériaux a montré qu'elles présentent très peu de variations sur l'intervalle des longueurs d'onde du visible [94], c'est à dire que leur spectre est quasiment plat. On parle alors de réflexion spéculaire neutre en supposant que la réflectance spéculaire de la surface élémentaire observée est indépendante de la longueur d'onde :

$$F_{spec}(\mathbf{x}, \lambda) = F_{spec}(\mathbf{x}). \quad (2.7)$$

Cette hypothèse est couramment admise puisque la plupart des modèles considère que l'information chromatique est portée par la seule réflexion diffuse (et non par la réflexion spéculaire).

#### Hypothèse 5 : Surface matte

Lorsqu'une surface est matte, il est courant d'utiliser le modèle Lambertien de formation de la couleur, mais certains auteurs préfèrent partir d'un modèle plus complexe (Shafer ou Kubelka-Munk) et suppose ensuite que la surface est matte. Cela revient simplement à annuler la réflexion spéculaire dans le modèle :

$$F_{spec}(\mathbf{x}, \lambda) = 0. \quad (2.8)$$

### 2.2.3 Hypothèses sur les sensibilités des capteurs de la caméra

#### Hypothèse 6 : Intégrale des courbes de sensibilité spectrale

Cette hypothèse suppose que les sensibilités spectrales  $k(\lambda)$ ,  $k = R, V, B$ , des capteurs de la caméra sont normalisées de sorte que les intégrales de chaque

courbe sur le spectre visible soient égales à une constante notées  $i_{RVB}$  :

$$\int_{380nm}^{780nm} R(\lambda) d\lambda = \int_{380nm}^{780nm} V(\lambda) d\lambda = \int_{380nm}^{780nm} B(\lambda) d\lambda = i_{RVB}. \quad (2.9)$$

### Hypothèse 7 : Bandes-passantes des capteurs

Cette hypothèse considère que les bandes passantes des sensibilités spectrales des capteurs  $k(\lambda)$ ,  $k = R, V, B$ , sont si étroites que chaque capteur associé à la sensibilité  $k(\lambda)$  n'est sensible qu'à une seule longueur d'onde notée  $\lambda_k$  [41] :

$$\int_{\lambda} k(\lambda) d\lambda = k(\lambda_k), \quad k = R, V, B. \quad (2.10)$$

Cette hypothèse, pourtant très utilisée, est rarement vérifiée en pratique. Finlayson *et al.* proposent donc d'affiner artificiellement les bandes-passantes des capteurs en appliquant une transformation affine sur les composantes  $RVB$  fournies par la caméra [29].

### Hypothèse 8 : Réponses des capteurs RVB transformées en composantes CIE 1964 XYZ

Pour obtenir les caractéristiques invariantes proposées par Geusebroek *et al.* [45], il est nécessaire de travailler dans l'espace colorimétrique CIE 1964 XYZ. Geusebroek *et al.* supposent que si, ni les capteurs de la caméra, ni l'éclairage ne sont connus, la transformation affine proposée par l'ITU [55] permet d'obtenir une assez bonne approximation des composantes XYZ à partir des composantes RVB fournies par ce caméra :

$$\begin{bmatrix} C^X(P) \\ C^Y(P) \\ C^Z(P) \end{bmatrix} = \begin{pmatrix} 0.62 & 0.11 & 0.19 \\ 0.3 & 0.56 & 0.05 \\ -0.01 & 0.03 & 1.11 \end{pmatrix} \begin{bmatrix} C^R(P) \\ C^V(P) \\ C^B(P) \end{bmatrix}. \quad (2.11)$$

Si les sensibilités spectrales des capteurs sont connues, l'approximation peut

être affinée.

## 2.2.4 Hypothèses sur les caractéristiques de l'illuminant

### Hypothèse 9 : Equation de Planck

Finlayson propose un modèle d'illuminant issu du modèle de Planck [35]. Ce modèle exprime la répartition spectrale relative d'énergie de l'illuminant qui atteint l'élément de surface à la position  $\mathbf{x}$  par :

$$E(\mathbf{x}, \lambda) = \frac{e(\mathbf{x})c_1}{\lambda^5 \left( \exp\left(\frac{c_2}{T(\mathbf{x})\lambda}\right) - 1 \right)}, \quad (2.12)$$

avec

$$\begin{cases} e(\mathbf{x}) : \text{l'intensité de l'illuminant en } \mathbf{x} \\ T(\mathbf{x}) : \text{la température de couleurs en Kelvin de l'illuminant en } \mathbf{x} \\ c_1 = 3.74183 \times 10^{-16} \text{ Wm}^2 \\ c_2 = 1.4388 \times 10^{-2} \text{ mK} \end{cases} \quad (2.13)$$

De plus, comme  $\lambda \in [10^{-7}; 10^{-6}]$  dans le visible et que  $T(\mathbf{x}) \in [10^3; 10^4]$ , Finlayson pose  $\exp\left(\frac{c_2}{T(\mathbf{x})\lambda}\right) \gg 1$ .

Par conséquent, il obtient le modèle d'illuminant suivant :

$$E(\mathbf{x}, \lambda) = \frac{e(\mathbf{x})c_1}{\lambda^5 \exp\left(\frac{c_2}{T(\mathbf{x})\lambda}\right)}. \quad (2.14)$$

### Hypothèse 10 : Couleur de l'éclairage constant localement

Gevers *et al.* [48] considèrent que l'illuminant  $E(\mathbf{x}, \lambda)$  peut se factoriser sous la forme d'un produit de deux termes. L'un, noté  $e(\mathbf{x})$ , dépend de la position  $\mathbf{x}$  de la surface élémentaire et va moduler l'intensité de l'éclairage. L'autre est une courbe spectrale  $E(\lambda)$  identique pour tous les éléments de surface observés par



des pixels d'un même voisinage :

$$E(\mathbf{x}, \lambda) = e(\mathbf{x}) \cdot E(\lambda). \quad (2.15)$$

Cette hypothèse revient donc à supposer que la couleur de l'éclairage est constante localement et que seule l'intensité varie en fonction de la position.

### **Hypothèse 11 : Illuminant constant localement**

Funt *et al.* [41] supposent que deux éléments de surface observés par deux pixels voisins dans une image sont éclairés de la même manière. Ainsi, si l'élément de surface à la position  $\mathbf{x}$  et projeté sur le pixel  $P$  est éclairé par un illuminant de répartition spectrale relative d'énergie  $E(\mathbf{x}, \lambda)$ , alors les éléments de surface projetés sur les pixels  $P_{vois}$  qui appartiennent au voisinage de  $P$  sont également éclairés par un illuminant de même répartition spectrale relative d'énergie. L'éclairage ne dépend donc pas de la position :

$$E(\mathbf{x}, \lambda) = E(\lambda). \quad (2.16)$$

Il n'y a pas de restriction sur l'étendue du voisinage considéré mais plus celle-ci est élevée, moins l'hypothèse aura de chance d'être vérifiée. Le voisinage  $3 \times 3$  est souvent retenu mais dans le cadre de la reconnaissance d'objets par description locale, il n'est pas rare de considérer que l'éclairage est constant pour toute la région locale extraite autour d'un point d'intérêt.

Un éclairage constant localement implique à la fois que le spectre de l'éclairage qui atteint deux surfaces élémentaires voisines est constant mais aussi que l'orientation de ces surfaces par rapport aux rayons incidents est constante. La validité de cette hypothèse implique donc que l'éclairage soit uniforme et que l'objet soit plan localement.

### **Hypothèse 12 : Illuminant blanc**

Cette hypothèse considère que la répartition spectrale relative d'énergie  $E(\mathbf{x}, \lambda)$  de l'illuminant qui éclaire l'élément de surface à la position  $\mathbf{x}$  est constante sur toutes les longueurs d'onde, l'intensité de cet illuminant n'étant pas forcément fixe d'une surface élémentaire à l'autre. Il s'agit d'un illuminant blanc, dont la répartition spectrale relative d'énergie est indépendante de la longueur d'onde :

$$E(\mathbf{x}, \lambda) = E(\mathbf{x}). \quad (2.17)$$

### **Hypothèse 13 : Couleur de l'éclairage connue**

Lors de la définition de caractéristiques couleur invariantes aux ombres et aux reflets, il peut parfois être nécessaire de connaître la couleur de l'éclairage [100]. L'hypothèse selon laquelle la couleur de l'éclairage qui illumine la scène est connue est très restrictive. Ceci dit, un certain nombre de méthodes permettent d'estimer cette couleur [36, 39].

## **2.2.5 Hypothèses sur le modèle de variations photométriques et radiométriques**

Dans ce paragraphe, nous allons présenter différentes hypothèses de variations photométriques et radiométriques, fondées sur différents modèles élaborés à partir de la fonction  $F_{ill}$  de l'équation (2.1). Nous rappelons que cette fonction modélise l'impact d'une variation photométrique ou radiométrique puisqu'elle permet le passage de la couleur d'un pixel  $P_{req}$  dans l'image requête  $\mathbf{I}_{req}$  à la couleur du pixel  $P_{cand}$  de l'image candidate  $\mathbf{I}_{cand}$ . Les deux images  $\mathbf{I}_{req}$  et  $\mathbf{I}_{cand}$  représentent le même objet et ont été acquises sous des conditions différentes. Les pixels  $P_{req}$  et  $P_{cand}$  représentent la même surface élémentaire de cet objet.

### Hypothèse 14 : Modèle diagonal

Certaines analyses expérimentales ont montré que les répartitions spectrales de la plupart des illuminants et que les réflectances spectrales de la plupart des matériaux peuvent être modélisées par des sommes de quelques fonctions spectrales de base [32, 56, 70] :

$$\beta(\mathbf{x}, \lambda) \approx \sum_i^{d\beta} \beta_i(\lambda) \alpha_i(\mathbf{x}), \quad (2.18)$$

et

$$E(\mathbf{x}, \lambda) \approx \sum_j^{dE} E_j(\lambda) \omega_j(\mathbf{x}), \quad (2.19)$$

où  $d\beta$  ( $dE$  resp.) est le nombre de fonctions de base  $\beta_i(\lambda)$  ( $E_j(\lambda)$ , resp.) et  $\alpha_i(\mathbf{x})$  ( $\omega_j(\mathbf{x})$  resp.) sont les poids correspondants.

Finalyson *et al.* [32, 33] ont montré que si  $d\beta = 3$  et  $dE = 2$  ou si  $d\beta = 2$  et  $dE = 3$ , un changement d'éclairage pouvait être modélisé par une simple matrice diagonale. Ainsi, si les changements de condition d'acquisition se limitent uniquement à un changement d'illuminant, le modèle diagonal suppose que la couleur  $\mathbf{C}(P_{cand}) = (C^R(P_{cand}), C^V(P_{cand}), C^B(P_{cand}))^t$  du pixel  $P_{cand}$  se déduit de la couleur  $\mathbf{C}(P_{req}) = (C^R(P_{req}), C^V(P_{req}), C^B(P_{req}))^t$  du pixel  $P_{req}$  grâce à la transformation  $F_{ill}$  définie par la matrice diagonale :

$$\mathbf{C}(P_{cand}) = \begin{pmatrix} a_R & 0 & 0 \\ 0 & a_V & 0 \\ 0 & 0 & a_B \end{pmatrix} \mathbf{C}(P_{req}). \quad (2.20)$$

Tous les modèles de variations photométriques et radiométriques résultant d'une transformation définie par une seule matrice diagonale, sont appelés modèles diagonaux [33]. Ils se distinguent les uns des autres uniquement par les valeurs et les dépendances des termes de la matrice diagonale [41, 48]. Le modèle diagonal peut être justifié, par exemple, en supposant que la surface est matte et

que les capteurs d'acquisition ont des bandes-passantes très étroites [41]. Si les bandes-passantes des capteurs sont larges, nous avons déjà précisé que Finlayson *et al.* ont proposé une méthode qui permet de les affiner artificiellement par une combinaison linéaire [29]. Cette transformation doit être adaptée à chaque système d'acquisition dont les courbes de sensibilité spectrales doivent être connues. Dans l'équation précédente, lorsque  $a_R = a_V = a_B$ , la variation se limite à un changement d'intensité de l'éclairage.

### Hypothèse 15 : Modèle diagonal et translation

Dans le modèle diagonal précédent, le terme de lumière ambiante introduit par Shafer [94] (terme  $L_a(\lambda)$  dans l'équation (2.5)) a été négligé. Aussi, Finlayson *et al.* [30] proposent de lever cette contrainte en ajoutant un terme de translation dans l'équation précédente. La transformation  $F_{III}$  est alors définie par deux matrices : une matrice 3x3 diagonale et une matrice 3x1 [98] :

$$\mathbf{C}(P_{cand}) = \begin{pmatrix} a_R & 0 & 0 \\ 0 & a_V & 0 \\ 0 & 0 & a_B \end{pmatrix} \mathbf{C}(P_{req}) + \begin{pmatrix} b_R \\ b_V \\ b_B \end{pmatrix}. \quad (2.21)$$

Finlayson *et al.* ont montré que la prise en compte de cette translation éventuelle dans des composantes couleurs permet de modéliser une plus large plage de variations des conditions d'acquisition [30].

### Hypothèse 16 : Transformation linéaire

Lorsque les sensibilités spectrales des capteurs sont inconnues et ne peuvent être affinées ou lorsque la propriété de faible dimension de l'illuminant et des réflectances spectrales ne sont pas validées [88, 105], la matrice diagonale devient

une matrice  $3 \times 3$  [25, 65, 96] :

$$\mathbf{C}(P_{cand}) = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \mathbf{C}(P_{req}). \quad (2.22)$$

Ce type de modèle considère qu'une composante  $C^k(P_{cand})$  dépend des 3 composantes du pixel  $P_{req}$  et non pas uniquement de la composante  $C^k(P_{req})$ .

Drew *et al.* [25] ont montré que ce modèle est théoriquement validé dans des conditions d'illumination très générales (plusieurs illuminants de couleurs, d'intensités et de directions différentes).

Van de Weijer *et al.* [100] estime également qu'une matrice  $3 \times 3$  est nécessaire pour modéliser une variation des conditions d'acquisition et notamment un changement de capteurs.

### Hypothèse 17 : Transformation affine

Lorsque l'on ajoute le terme de lumière ambiante au modèle précédent, les variations photométriques et radiométriques peuvent être modélisées par la transformation suivante [77] :

$$\mathbf{C}(P_{cand}) = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \mathbf{C}(P_{req}) + \begin{pmatrix} j \\ k \\ l \end{pmatrix}. \quad (2.23)$$

De tous ces modèles, nous pouvons déduire que la majorité des changements d'éclairage et même de capteurs peut être modélisée par l'application d'une transformation affine dans l'espace couleur, même si le phénomène de métamérisme le contredit. Cependant, il est à noter qu'il n'y a pas de consensus sur le nombre de degrés de liberté de cette transformation : 3 pour le modèle diagonal recommandé par [33, 41, 48], 6 pour le modèle diagonal et translation recommandé

par [30, 51, 77], 9 pour le modèle linéaire recommandé par [25] et 12 pour le modèle affine recommandé par [75]. Dans le contexte de la description locale, ces modèles sont utilisés dans le seul but de normaliser les couleurs des pixels afin de les rendre invariants aux conditions d'acquisition. Il est démontré que, plus nous augmentons le nombre de degrés de liberté, plus le degré d'invariance est élevé avec le risque de diminuer en parallèle le pouvoir discriminant. Un compromis entre pouvoir discriminant et degré d'invariance doit donc être pris en compte dans le choix du modèle.

### **Hypothèse 18 : Fonctions monotones croissantes**

Finlayson [28] considère indépendamment chaque composante couleur et suppose que, le niveau  $C^k(P_{cand})$ ,  $k = R, V, B$ , du pixel  $P_{cand}$  est déduit du niveau  $C^k(P_{req})$  du pixel correspondant  $P_{req}$  à l'aide d'une fonction strictement croissante  $f^k$  :

$$C^k(P_{cand}) = f^k(C^k(P_{req})), k = R, V, B. \quad (2.24)$$

Nous rappelons qu'une fonction  $f^k$  est strictement croissante si  $a > b \Rightarrow f^k(a) > f^k(b)$ . Les trois fonctions monotones croissantes  $f^k$ ,  $k = R, V, B$ , ne sont pas a priori des fonctions linéaires, comme c'est le cas pour un modèle diagonal.

Toutes ces hypothèses forment la base des principales caractéristiques couleur invariantes que nous proposons de présenter dans le paragraphe suivant.

## **2.3 Les caractéristiques couleur invariantes**

Dans ce paragraphe, nous proposons de présenter les principales caractéristiques couleur invariantes qui sont utilisées dans le contexte de la reconnaissance d'objets. L'objectif n'étant pas d'avoir une présentation exhaustive des caractéristiques couleur invariantes mais de comprendre comment celles-ci sont définies et sur quelle(s) hypothèse(s) de base elles sont construites. Nous proposons de les classer en 3 catégories. La première catégorie regroupe toutes les caractéristiques

invariantes qui sont obtenues en calculant les rapports intra ou inter composantes couleur d'un même pixel ou entre les composantes couleur de pixels voisins. La seconde catégorie regroupe les caractéristiques invariantes qui exploitent la distribution couleur des pixels qui constituent la région locale à caractériser. Les caractéristiques invariantes de la troisième catégorie sont les dérivées spectrales et/ou spatiales des composantes couleur originales.

### 2.3.1 Les rapports inter et intra composantes couleur

#### (a) Rapports entre composantes de pixels voisins

- *L'approche de Funt et al. [41]*

Funt *et al.* proposent un modèle diagonal de changement d'éclairage fondé sur les trois hypothèses suivantes :

- Modèle Lambertien de formation de la couleur (équation (2.6)) :  

$$C_{Lambert}(\mathbf{x}, \lambda) = \beta(\mathbf{x}, \lambda) E(\mathbf{x}, \lambda).$$
- Bandes-passantes des capteurs étroites (équation (2.10)) :  $\int_{\lambda} k(\lambda) d\lambda = k(\lambda_k).$
- Illuminant constant sur un voisinage  $3 \times 3$  et égal à  $E(V_{3 \times 3}(P), \lambda)$  : équation (2.16).

Ainsi, en reprenant les notations du paragraphe précédent, les composantes couleur  $C^k(P)$ ,  $k = R, V, B$ , du pixel  $P$  qui observe la surface élémentaire à la position  $\mathbf{x}$  sont estimées par :

$$\begin{aligned} C^k(P) &= \int_{\lambda} \beta(\mathbf{x}, \lambda) \cdot E(V_{3 \times 3}(P), \lambda) \cdot k(\lambda) d\lambda \\ &= \beta(\mathbf{x}, \lambda_k) \cdot E(V_{3 \times 3}(P), \lambda_k) \cdot k(\lambda_k). \end{aligned} \tag{2.25}$$

De la même manière, les composantes  $C^k(P_{vois})$ ,  $k = R, V, B$ , du pixel  $P_{vois}$

appartenant au voisinage  $3 \times 3$  du pixel  $P$  et qui observe sur la surface élémentaire à la position  $\mathbf{x}_{\text{vois}}$  sont estimées par :

$$C^k(P_{\text{vois}}) = \beta(\mathbf{x}_{\text{vois}}, \lambda_k) \cdot E(V_{3 \times 3}(P_{\text{vois}}), \lambda_k) \cdot k(\lambda_k). \quad (2.26)$$

Des équations (2.25) et (2.26), nous en déduisons que le rapport suivant :

$$\frac{C^k(P)}{C^k(P_{\text{vois}})} = \frac{\beta(\mathbf{x}, \lambda_k)}{\beta(\mathbf{x}_{\text{vois}}, \lambda_k)}, \quad (2.27)$$

ne dépend que des réflectances spectrales des surfaces élémentaires observées et des sensibilités spectrales des capteurs (de par leurs longueurs d'ondes dominantes  $\lambda_k$ ).

Ainsi, si la surface est lambertienne, que les bandes passantes des capteurs sont étroites et que l'éclairage est uniforme localement, Funt *et al.* ont montré que la division du niveau d'un pixel par le niveau d'un pixel voisin était invariante à la couleur de l'éclairage, à l'intensité de l'éclairage et à l'angle d'observation (conséquence inhérente à l'hypothèse d'une surface lambertienne).

Funt *et al.* [41] proposent donc d'exprimer les caractéristiques couleur invariantes  $(X^1(P), X^2(P), X^3(P))^t$  du pixel  $P$  de la manière suivante :

$$\begin{cases} X^1(P) = \log(C^R(P_{\text{vois}})) - \log(C^R(P)), \\ X^2(P) = \log(C^V(P_{\text{vois}})) - \log(C^V(P)), \\ X^3(P) = \log(C^B(P_{\text{vois}})) - \log(C^B(P)), \end{cases} \quad (2.28)$$

où  $P_{\text{vois}}$  est l'un des pixels du voisinage du pixel  $P$  dans l'image considérée.

Partant d'un modèle équivalent, Nayar *et al.* [86] et Koubaroulis *et al.* [62] ont aussi proposé des rapports entre niveaux de composantes couleur de pixels voisins.



Si on reprend les hypothèses de Funt *et al.* et on étend l'hypothèse de l'uniformité de l'éclairage à toute la région extraite autour d'un point d'intérêt, d'autres approches deviennent alors possible. Ainsi, si on s'inspire des travaux de Gershon *et al.* [42], nous pouvons diviser le niveau d'un pixel par la moyenne des niveaux de tous les pixels de la région extraite et appliquer ainsi la "normalisation du monde gris" locale [84]. De même, partant des mêmes hypothèses, certains auteurs préfèrent normaliser les niveaux des pixels, non pas par le niveau moyen dans le voisinage, mais par le niveau maximum obtenu par tous les pixels de ce voisinage [22, 64]. Finlayson *et al.* [38] ont d'ailleurs proposé une généralisation de ces normalisations par des mesures statistiques et ont montré que ces dernières normalisations ne sont que des cas particuliers d'une normalisation fondée sur les normes de Minkowski d'ordre 1 à l'infini. Dans le cas particulier de l'estimation de l'illuminant, ils ont montré que l'ordre 6 fournit les meilleurs résultats. Dans le cas de la reconnaissance d'objets par description locale, ces normalisations n'ont pas été testées mais reposent sur les mêmes hypothèses que celles définies plus haut et pourraient donc être appliquées.

Partant toujours des mêmes hypothèses et considérant que l'éclairage est uniforme sur toute la région extraite autour d'un point caractéristique, Finlayson considère un espace à  $N_{pix}$  dimensions contenant 3 vecteurs  $\mathbf{VEC}^R$ ,  $\mathbf{VEC}^V$  et  $\mathbf{VEC}^B$ , chacun associé respectivement à une composante couleur  $R$ ,  $V$  et  $B$ . Le nombre  $N_{pix}$  correspond au nombre de pixels dans la région considérée. Les coordonnées  $VEC_1^k, VEC_2^k, \dots, VEC_{N_{pix}}^k$  des vecteurs  $\mathbf{VEC}^k$ ,  $k = R, V, B$ , sont les niveaux de composante  $C^k(P)$  des pixels  $P$  de la région. Les pixels sont considérés dans le même ordre pour les trois composantes, de sorte que les valeurs  $VEC_j^R, VEC_j^V$  et  $VEC_j^B$ ,  $j = 1 \dots N_{pix}$ , correspondent aux coordonnées du  $j^{ème}$  pixel considéré dans la région. Finlayson constate que si les hypothèses et contraintes présentées ci-avant sont respectées, les angles  $angle^{kk'}$ ,  $k, k' = R, V, B$  formés par les vecteurs  $\mathbf{VEC}^k$  et  $\mathbf{VEC}^{k'}$  sont

insensibles aux changements d'illuminant [31]. En effet, un changement d'illuminant, dans le cadre de ces hypothèses, provoque une multiplication des niveaux de tous les pixels par une même constante. Cela entraîne donc une modification de la norme de chaque vecteur  $\mathbf{VEC}^k$  associé à chaque composante couleur  $k$ , sans en modifier la direction. Finlayson propose donc de considérer les 3 valeurs qui représentent les angles entre les trois vecteurs couleur comme des caractéristiques couleur invariantes.

- *Les caractéristiques  $m_1, m_2, m_3$  de Gevers et al. [48]*

Gevers *et al.* proposent des caractéristiques invariantes notées  $\{m_1, m_2, m_3\}$ , lesquelles reposent sur les 4 hypothèses suivantes [48] :

- Modèle dichromatique de Shafer (équation (2.4)) :  

$$C_{Shafer}(\mathbf{x}, \lambda) = m_{diff}(\theta) \cdot \beta(\mathbf{x}, \lambda) : E(\mathbf{x}, \lambda) + m_{spec}(\theta, \alpha) \cdot F_{spec}(\mathbf{x}, \lambda) \cdot E(\mathbf{x}, \lambda)$$
- Surface matte (équation (2.8)) :  $F_{spec}(\mathbf{x}, \lambda) = 0$ .
- Bandes-passantes des capteurs étroites (équation (2.10)) :  $\int_{\lambda} k(\lambda) d\lambda = k(\lambda_k)$ .
- Couleur de l'éclairage constante localement (équation (2.15)) :  $E(\mathbf{x}, \lambda) = e(\mathbf{x}) \cdot E(\lambda)$ .

Ainsi, les composantes couleur  $C^k(P)$ ,  $k = R, V, B$ , du pixel  $P$  qui observe sur la surface élémentaire à la position  $\mathbf{x}$  sont estimées par :

$$C^k(P) = m_{diff}(\theta) \cdot \beta(\mathbf{x}, \lambda_k) \cdot e(\mathbf{x}) \cdot E(\lambda_k) \cdot k(\lambda_k). \quad (2.29)$$

De la même manière, les composantes  $C^k(P_{vois})$ ,  $k = R, V, B$ , du pixel  $P_{vois}$  appartenant au voisinage  $3 \times 3$  du pixel  $P$  et qui observe sur la surface élémentaire à la position  $\mathbf{x}_{vois}$  sont estimées par :

$$C^k(P_{vois}) = m_{diff}(\theta_{vois}) \cdot \beta(\mathbf{x}_{vois}, \lambda_k) \cdot e(\mathbf{x}_{vois}) \cdot E(\lambda_k) \cdot k(\lambda_k), \quad (2.30)$$

où  $\theta_{vois}$  représente la direction de l'éclairage par rapport à la normale de la surface en  $\mathbf{x}_{vois}$ . Cette direction peut être différente de  $\theta$  qui représente la direction de l'éclairage par rapport à la normale de la surface en  $\mathbf{x}$ . Ce paramètre permet de prendre en compte les variations d'orientation par rapport à la surface.

Si nous considérons une autre composante couleur  $k' \neq k$ , nous pouvons calculer les composantes couleur des pixels  $P$  et  $P_{vois}$  :

$$C^{k'}(P) = m_{diff}(\theta) \cdot \beta(\mathbf{x}, \lambda'_k) \cdot e(\mathbf{x}) \cdot E(\lambda'_k) \cdot k'(\lambda'_k) \quad (2.31)$$

et

$$C^{k'}(P_{vois}) = m_{diff}(\theta_{vois}) \cdot \beta(\mathbf{x}_{vois}, \lambda'_k) \cdot e(\mathbf{x}_{vois}) \cdot E(\lambda'_k) \cdot k'(\lambda'_k) \quad (2.32)$$

En reprenant les 4 équations précédentes, il est facile de montrer que le rapport suivant ne dépend que des réflectances spectrales des surfaces élémentaires considérées et des sensibilités spectrales des capteurs (de par leurs longueurs d'ondes dominantes  $\lambda_k$ ) :

$$\frac{C^k(P)C^{k'}(P_{vois})}{C^k(P_{vois})C^{k'}(P)} = \frac{\beta(\mathbf{x}, \lambda_k)\beta(\mathbf{x}(P_{vois}), \lambda'_k)}{\beta(\mathbf{x}(P_{vois}), \lambda_k)\beta(\mathbf{x}, \lambda'_k)}. \quad (2.33)$$

Ainsi, pour une surface matte, lorsque les bandes-passantes des capteurs d'acquisition sont très étroites et que la couleur de l'éclairage est constante localement, Gevers *et al.* ont montré que le rapport croisé entre les niveaux de deux pixels voisins pour deux composantes différentes était insensible à la couleur, à l'intensité et à la direction de l'éclairage ainsi qu'à l'angle d'observation.

Gevers *et al.* proposent donc d'exprimer les caractéristiques couleur inva-

riantes  $(X^1(P), X^2(P), X^3(P))$  du pixel  $P$  de la manière suivante [48] :

$$\begin{cases} X^1(P) = \frac{C^R(P)C^V(P_{vois})}{C^R(P_{vois})C^V(P)}, \\ X^2(P) = \frac{C^R(P)C^B(P_{vois})}{C^R(P_{vois})C^B(P)}, \\ X^3(P) = \frac{C^V(P)C^B(P_{vois})}{C^V(P_{vois})C^B(P)}, \end{cases} \quad (2.34)$$

où  $P_{vois}$  est l'un des pixels du voisinage du pixel  $P$  dans l'image considérée.

### (b) Rapports entre composantes d'un même pixel

- Les caractéristiques  $l_1, l_2, l_3$  de Gevers *et al.* [48]

Gevers *et al.* proposent ensuite des caractéristiques invariantes notées  $\{l_1, l_2, l_3\}$  et reposant sur les quatre hypothèses suivantes :

- Modèle de Shafer de formation de la couleur (équation (2.4)) :  $C_{Shafer}(\mathbf{x}, \lambda) = m_{diff}(\theta) \cdot \beta(\mathbf{x}, \lambda) \cdot E(\mathbf{x}, \lambda) + m_{spec}(\theta, \alpha) \cdot F_{spec}(\mathbf{x}, \lambda) \cdot E(\mathbf{x}, \lambda)$ .
- Réflexion spéculaire neutre (équation (2.7)) :  $F_{spec}(\mathbf{x}, \lambda) = F_{spec}(\mathbf{x})$ .
- Intégrales des courbes de sensibilité spectrales des capteurs identiques (équation (2.9)) :  $\int_{380nm}^{780nm} R(\lambda) d\lambda = \int_{380nm}^{780nm} V(\lambda) d\lambda = \int_{380nm}^{780nm} B(\lambda) d\lambda = i_{RVB}$ .
- Illuminant blanc (équation (2.17)) :  $E(\mathbf{x}, \lambda) = E(\mathbf{x})$ .

Ainsi, les composantes couleur  $C^k(P)$ ,  $k = R, V, B$ , du pixel  $P$  qui observe la surface élémentaire à la position  $\mathbf{x}$  sont estimées par :

$$\begin{aligned} C^k(P) = & m_{diff}(\theta) \cdot E(\mathbf{x}) \int_{\lambda} \beta(\mathbf{x}, \lambda) \cdot k(\lambda) d\lambda \\ & + m_{spec}(\theta, \alpha) \cdot F_{spec}(\mathbf{x}) \cdot E(\mathbf{x}) i_{RVB}. \end{aligned} \quad (2.35)$$

De la même manière, en considérant les trois composantes différentes  $k, k'$  et  $k''$ , le rapport de différences suivant ne dépend que de la réflectance spec-

trale de la surface élémentaire considérée et des sensibilités des capteurs :

$$\frac{C^k(P) - C^{k'}(P)}{C^k(P) - C^{k''}(P)} = \frac{\int_{\lambda} \beta(\mathbf{x}, \lambda) \cdot k(\lambda) d\lambda - \int_{\lambda} \beta(\mathbf{x}, \lambda) \cdot k'(\lambda) d\lambda}{\int_{\lambda} \beta(\mathbf{x}, \lambda) \cdot k(\lambda) d\lambda - \int_{\lambda} \beta(\mathbf{x}, \lambda) \cdot k''(\lambda) d\lambda} \quad (2.36)$$

Ainsi, en cas de réflexion spéculaire neutre, lorsque les intégrales des capteurs sont identiques et que l'illuminant est blanc, Gevers *et al.* ont montré que, pour un même pixel, le rapport entre les différences des niveaux de composantes différentes était insensible à l'intensité de l'éclairage, à la direction de l'éclairage, à l'angle d'observation et aux reflets (réflexion spéculaire).

Gevers *et al.* [48] proposent donc d'exprimer les caractéristiques couleur invariantes  $(X^1(P), X^2(P), X^3(P))^t$  du pixel  $P$  de la manière suivante :

$$\begin{cases} X^1(P) = \frac{(C^R(P) - C^V(P))^2}{(C^R(P) - C^V(P))^2 + (C^R(P) - C^B(P))^2 + (C^V(P) - C^B(P))^2}, \\ X^2(P) = \frac{(C^R(P) - C^B(P))^2}{(C^R(P) - C^V(P))^2 + (C^R(P) - C^B(P))^2 + (C^V(P) - C^B(P))^2}, \\ X^3(P) = \frac{(C^V(P) - C^B(P))^2}{(C^R(P) - C^V(P))^2 + (C^R(P) - C^B(P))^2 + (C^V(P) - C^B(P))^2}. \end{cases} \quad (2.37)$$

Cette propriété peut être généralisée à toutes les caractéristiques couleur fondées sur le rapport entre les différences des niveaux de composantes différentes d'un même pixel. C'est le cas, par exemple, de la teinte  $H$  définie par l'équation  $H(P) = \arctan\left(\frac{\sqrt{3}(C^V(P) - C^B(P))}{(C^R(P) - C^V(P)) + (C^R(P) - C^B(P))}\right)$ .

- Les composantes  $c_1, c_2, c_3$  de Gevers *et al.* [48]

Gevers *et al.* proposent enfin des caractéristiques invariantes notées  $\{c_1, c_2, c_3\}$ , lesquelles reposent sur les 3 hypothèses suivantes :

- Modèle dichromatique de Shafer (équation (2.4)) :

$$C_{Shafer}(\mathbf{x}, \lambda) = m_{diff}(\theta) \cdot \beta(\mathbf{x}, \lambda) \cdot E(\mathbf{x}, \lambda) + m_{spec}(\theta, \alpha) \cdot F_{spec}(\mathbf{x}, \lambda) \cdot E(\mathbf{x}, \lambda)$$

- Surface matte (équation (2.8)) :  $F_{spec}(\mathbf{x}, \lambda) = 0$ .

– Illuminant blanc (équation (2.17)) :  $E(\mathbf{x}, \lambda) = E(\mathbf{x})$ .

Ainsi, les composantes couleur  $C^k(P)$ ,  $k = R, V, B$ , du pixel  $P$  qui observé sur la surface élémentaire à la position  $\mathbf{x}$  sont estimées par :

$$C^k(P) = m_{diff}(\theta) \cdot E(\mathbf{x}) \cdot \int_{\lambda} \beta(\mathbf{x}, \lambda) \cdot k(\lambda) d\lambda. \quad (2.38)$$

De la même manière, si l'on considère la composante couleur  $C^{k'}$ , avec  $k' \neq k$ , nous obtenons :

$$C^{k'}(P) = m_{diff}(\theta) \cdot E(\mathbf{x}) \cdot \int_{\lambda} \beta(\mathbf{x}, \lambda) \cdot k'(\lambda) d\lambda. \quad (2.39)$$

Le rapport entre ces deux composantes couleur ne dépend donc que de la réflectance spectrale de la surface élémentaire et des sensibilités des capteurs :

$$\frac{C^k(P)}{C^{k'}(P)} = \frac{\int_{\lambda} \beta(\mathbf{x}, \lambda) \cdot k(\lambda) d\lambda}{\int_{\lambda} \beta(\mathbf{x}, \lambda) \cdot k'(\lambda) d\lambda}. \quad (2.40)$$

Ainsi, lorsque la surface est matte et que l'éclairage est blanc, Gevers *et al.* ont montré que, pour un même pixel, le rapport entre le niveau d'une composante et le niveau d'une autre composante était insensible à l'intensité de l'éclairage, à la direction de l'éclairage et à l'angle d'observation.

Gevers *et al.* proposent donc d'exprimer les caractéristiques couleur invariantes  $(X^1(P), X^2(P), X^3(P))^t$  du pixel  $P$  de la manière suivante [48] :

$$\begin{cases} X^1(P) = \arctan\left(\frac{C^R(P)}{\max(C^V(P), C^B(P))}\right), \\ X^2(P) = \arctan\left(\frac{C^V(P)}{\max(C^R(P), C^B(P))}\right), \\ X^3(P) = \arctan\left(\frac{C^B(P)}{\max(C^R(P), C^V(P))}\right). \end{cases} \quad (2.41)$$

Ici, la fonction  $\max()$  retourne bien le maximum des niveaux d'un même pixel pour deux composantes données et non le maximum pour tous les

pixels de la région.

Cette propriété peut être généralisée à toutes les caractéristiques couleur fondées sur le rapport entre les niveaux de différentes composantes d'un même pixel. C'est le cas, par exemple, des composantes normalisées  $C^r(P) = \frac{C^R(P)}{C^R(P)+C^V(P)+C^B(P)}$ ,  $C^v(P) = \frac{C^V(P)}{C^R(P)+C^V(P)+C^B(P)}$  et  $C^b(P) = \frac{C^B(P)}{C^R(P)+C^V(P)+C^B(P)}$ . Finlayson *et al.* avaient bien noté que deux modèles différents étaient nécessaires pour obtenir des caractéristiques couleur invariantes à la fois à la couleur et à la direction de l'éclairage [37]. En effet, les caractéristiques proposées par Funt *et al.* [41] normalisent les composantes couleur par rapport à la couleur de l'éclairage alors que les composantes normalisées  $\{C^r(P), C^v(P), C^b(P)\}$  sont invariantes à la direction de l'éclairage, et donc à l'orientation de la surface considérée. Ils proposent donc une normalisation itérative des composantes couleur des pixels convergeant rapidement vers des caractéristiques couleur invariantes à la fois à la couleur et à la direction de l'éclairage [37]. Cette normalisation consiste simplement à alterner la normalisation proposée par Funt *et al.* [41] et la normalisation qui fournit les composantes couleur normalisées  $\{C^r(P), C^v(P), C^b(P)\}$ .

- *L'approche de Finlayson et al. [35]*

Finlayson *et al.* proposent des caractéristiques invariantes qui reposent sur les trois hypothèses suivantes :

- Modèle Lambertien de formation de la couleur (équation (2.6)) :

$$C_{Lambert}(\mathbf{x}, \lambda) = \beta(\mathbf{x}, \lambda) E(\mathbf{x}, \lambda).$$

- Bandes-passantes des capteurs étroites (équation (2.10)) :

$$\int_{\lambda} k(\lambda) d\lambda = k(\lambda_k).$$

- Equation de Planck pour l'illuminant (équation (2.14)) :

$$E(\mathbf{x}, \lambda) = \frac{e(\mathbf{x})c_1}{\lambda^5 \exp(\frac{c_2}{T(\mathbf{x})\lambda})}.$$

## CHAPITRE 2. LES DESCRIPTEURS COULEUR INVARIANTS DE LA LITTÉRATURE

---

Ainsi, en reprenant les notations précédentes, les composantes couleur  $C^k(P)$ ,  $k = R, V, B$ , du pixel  $P$  qui observe la surface élémentaire à la position  $\mathbf{x}$  sont estimées par :

$$C^k(P) = \beta(\mathbf{x}, \lambda_k) \cdot \frac{e(\mathbf{x})c_1}{\lambda_k^5 \exp(\frac{c_2}{T(\mathbf{x})\lambda_k})} \cdot k(\lambda_k). \quad (2.42)$$

En prenant le logarithme népérien de l'équation précédente, nous obtenons :

$$\ln(C^k(P)) = \ln(e(\mathbf{x})) + \ln\left(\frac{\beta(\mathbf{x}, \lambda_k)c_1 k(\lambda_k)}{\lambda_k^5}\right) - \frac{c_2}{T(\mathbf{x})\lambda_k}. \quad (2.43)$$

Cette équation peut être simplifiée de la manière suivante :

$$\ln(C^k(P)) = \text{Int}(\mathbf{x}) + \text{Ref}_k(\mathbf{x}) + T^{-1}(\mathbf{x})L_k, \quad (2.44)$$

où :

- $\text{Int}(\mathbf{x}) = \ln(e(\mathbf{x}))$  dépend de l'intensité de l'éclairage,
- $\text{Ref}_k(\mathbf{x}) = \ln\left(\frac{\beta(\mathbf{x}, \lambda_k)c_1 k(\lambda_k)}{\lambda_k^5}\right)$  dépend des propriétés de réflectance de la surface  $\mathbf{x}$  ainsi que des longueurs d'ondes dominantes  $\lambda_k$  des capteurs d'acquisition,
- $T^{-1}(\mathbf{x})L_k = -\frac{c_2}{T(\mathbf{x})\lambda_k}$  dépend de la couleur de l'éclairage et des capteurs d'acquisition.

Si nous considérons maintenant une seconde composante  $k' \neq k$ , nous pouvons calculer la différence des logarithmes qui ne dépend pas de l'intensité



de l'éclairage :

$$\begin{aligned} \ln(C^k(P)) - \ln(C^{k'}(P)) &= \ln\left(\frac{C^k(P)}{C^{k'}(P)}\right) = Ref_k(\mathbf{x}) - Ref_{k'}(\mathbf{x}) \\ &\quad + T^{-1}(\mathbf{x})(L_k - L_{k'}). \end{aligned} \quad (2.45)$$

De la même manière, si nous considérons les composantes  $k''$  et  $k'$ , nous obtenons :

$$\ln\left(\frac{C^{k''}(P)}{C^{k'}(P)}\right) = Ref_{k''}(\mathbf{x}) - Ref_{k'}(\mathbf{x}) + T^{-1}(\mathbf{x})(L_{k''} - L_{k'}). \quad (2.46)$$

En observant les deux équations précédentes, nous remarquons que lorsque la couleur de l'éclairage  $T(\mathbf{x})$  évolue pour une surface  $\mathbf{x}$  observée par un pixel  $P$ , les points correspondant dans l'espace  $(\ln(\frac{C^k(P)}{C^{k'}(P)}), \ln(\frac{C^{k''}(P)}{C^{k'}(P)}))$  se déplacent le long d'une ligne dont la direction dépend des capteurs d'acquisition. Par conséquent lorsque les caractéristiques des capteurs d'acquisition sont connus, Finlayson *et al.* proposent de calculer cette direction et de projeter tous les points de l'espace  $(\ln(\frac{C^V(P)}{C^R(P)}), \ln(\frac{C^B(P)}{C^R(P)}))$  sur une droite orthogonale à cette direction. Ils montrent que les coordonnées ainsi obtenues ne dépendent ni de l'intensité, ni de la couleur de l'éclairage. Ils montrent aussi que les ombres peuvent être éliminées des images grâce à cette projection. La différence avec les caractéristiques précédentes est qu'il n'y a qu'une caractéristique couleur invariante par pixel et non trois.

Faisant suite à ces travaux, Finlayson *et al.* ont montré qu'il n'était pas nécessaire de connaître les caractéristiques des capteurs d'acquisition pour définir la direction de la droite sur laquelle sont projetés les points [34]. En effet, ils ont montré que cette direction est celle qui minimise l'entropie de l'image invariante résultante.

### 2.3.2 Les transformations fondées sur une analyse des distributions colorimétriques

#### (a) Exploitation des mesures de rangs des pixels par Finlayson *et al.* [28]

Finlayson *et al.* [28] proposent un modèle fondé sur l'hypothèse qu'un changement d'illuminant peut être modélisé par l'application d'une fonction strictement croissante sur chaque composante (équation (2.24)).

Ils proposent alors de trier les pixels d'une région donnée de l'image dans l'ordre croissant de leurs niveaux pour chaque composante. La mesure de rang colorimétrique  $Rc^k(P)$  du pixel  $P$  pour la composante  $k$  est le rapport entre le nombre de pixels caractérisés par un niveau inférieur ou égal à celui de  $P$  et le nombre total de pixels dans cette région. Cette mesure de rang colorimétrique est donc définie de telle sorte qu'elle soit proche de 0 pour les premiers pixels (niveaux les plus faibles) et égale à 1 pour les derniers pixels (niveaux les plus élevés).

Considérons deux pixels  $P_1$  et  $P_2$  dans une région extraite autour d'un point d'intérêt. Si les composantes couleur de ces pixels sont telles que  $C^k(P_1) > C^k(P_2)$  alors, après changement d'illuminant, nous aurons  $f^k(C^k(P_1)) > f^k(C^k(P_2))$  puisque le changement d'illuminant est modélisé par une fonction strictement croissante  $f^k$  dans chaque composante. Autrement dit, si l'hypothèse de Finlayson *et al.* est vérifiée, l'ordre des niveaux des pixels n'est pas modifié en cas de changement d'illuminant. Par conséquent, les mesures de rangs de chaque pixel sont conservées en cas de variation d'éclairage. Finlayson *et al.* proposent donc de caractériser les pixels d'une région, non pas par leurs niveaux de composantes couleur  $C^k(P)$  mais par leurs mesures de rangs colorimétriques  $Rc^k(P)$ ,  $k=R, G, B$ , dans cette région. Ceci revient à effectuer indépendamment une égalisation des histogrammes de chaque composante de la région. Finlayson *et al.* ont montré que cette

normalisation fournit des caractéristiques couleur invariantes aux capteurs d'acquisition et à la couleur de l'éclairage.

### **(b) Normalisations fondées sur une transformation de la distribution colorimétrique**

Certains auteurs [49, 65] proposent de normaliser la distribution des couleurs d'une image (ou d'une région de l'image) dans l'espace couleur  $RVB$  pour en déduire des caractéristiques invariantes à l'éclairage.

Ainsi, Lenz *et al.* [65] supposent qu'un changement d'éclairage peut être modélisé par une transformation affine associée à une matrice  $3 \times 3$  (cf. Equation (2.24)). Ils calculent alors l'impact de cette transformation sur la distribution des couleurs des pixels dans l'espace couleur  $RVB$  et en déduisent une normalisation dans l'espace couleur qui est invariante à cette transformation affine. Pour cela, dans un premier temps, la distribution est normalisée de telle sorte que sa matrice des moments d'ordre 2 soit égale à la matrice unité. Ensuite, une rotation est appliquée de sorte que les moments d'ordre 3 soient normalisés.

Dans le même esprit, Healey *et al.* avaient supposé qu'un changement d'éclairage pouvait être modélisé par l'application d'une transformation affine  $3 \times 3$  à la distribution des couleurs dans l'espace  $RVB$  [49]. Ils ont ensuite proposé de normaliser la distribution couleur de telle sorte que les valeurs propres des matrices de moments soient invariantes aux variations de la couleur de l'éclairage.

Finlayson *et al.* [28] ont montré que le modèle diagonal associé à une translation reflète (cf. équation (2.21)) bien les conséquences d'un changement d'illuminant. Dans ce cas, la normalisation de la distribution colorimétrique qui consiste à imposer une moyenne nulle et un écart-type à 1 permet de rendre les composantes des pixels invariantes à ce type de changement

d'illuminant. Les caractéristiques couleur invariantes associées sont définies par les équations suivantes :

$$\begin{cases} X^1(P) = \frac{C^R(P) - \mu(C^R(P_i))}{\sigma(C^R(P_i))}, \\ X^2(P) = \frac{C^V(P) - \mu(C^V(P_i))}{\sigma(C^V(P_i))}, \\ X^3(P) = \frac{C^B(P) - \mu(C^B(P_i))}{\sigma(C^B(P_i))}, \end{cases} \quad (2.47)$$

où  $\mu(C^k(P_i))$  et  $\sigma(C^k(P_i))$  ( $k = R, V, B$ ) représentent respectivement la moyenne et l'écart-type des pixels  $P_i$  qui appartiennent à la région considérée.

Il est facile de montrer que la soustraction par la moyenne permet de supprimer la translation provoquée par le changement d'illuminant alors que la division par l'écart-type permet de supprimer les coefficients de la matrice diagonale.

Ces caractéristiques sont invariantes aux variations de l'intensité et de la couleur de l'illuminant, ainsi qu'aux variations liées au terme de lumière ambiante (cf. page 26, équation (2.5)).

### (c) Les moments couleur invariants de Mindru *et al.* [77]

Les moments de la distribution couleur ne prennent pas en compte l'information spatiale des pixels dans les images. Mindru *et al.* ont donc proposé d'utiliser plutôt les moments couleur généralisés qui pallient ce problème et qui sont définis de la manière suivante :

$$M_{pq}^{abc} = \int_y \int_x x^p y^q C^R(P_{xy})^a C^V(P_{xy})^b C^B(P_{xy})^c dx dy, \quad (2.48)$$

où  $(x, y)$  est la position du pixel  $P_{xy}$  dans l'image considérée.

Ainsi,  $M_{pq}^{abc}$  est appelé moment couleur généralisé d'ordre  $p + q$  et de degré  $a + b + c$ . Mindru *et al.* proposent alors de considérer les moments couleur généralisés d'ordre inférieur ou égale à 1 et de degré inférieur ou égal à

2. Ils considèrent ensuite 3 modèles de variations photométriques et radiométriques : le modèle diagonal (cf. equation (2.20)), le modèle diagonal et translation (cf. equation (2.21)) et la transformation affine et translation (cf. equation (2.23)) et montrent que dans chaque cas, on peut trouver plusieurs combinaisons de ces moments généralisés qui sont invariants à cette transformation. Étant donné que les positions des pixels dans l'image sont aussi prises en compte, ils proposent aussi des combinaisons qui sont invariantes à diverses transformations géométriques.

### 2.3.3 Les dérivées invariantes

#### (a) Les dérivées spectrales de Geusebroek *et al.* [45]

Geusebroek *et al.* ont proposé des caractéristiques couleur invariantes fondées sur des dérivées spectrales des réflectances des surfaces élémentaires. Pour cela, ils ont posé trois hypothèses communes à toutes leurs caractéristiques couleur invariantes :

- Modèle de Kubelka-Munk de formation de la couleur (équation (2.3)) :  

$$C_{Kub-Munk}(\mathbf{x}, \lambda) = (1 - F_{spec}(\mathbf{x}, \lambda))^2 \cdot \beta(\mathbf{x}, \lambda) \cdot E(\mathbf{x}, \lambda) + F_{spec}(\mathbf{x}, \lambda) \cdot E(\mathbf{x}, \lambda).$$
- Le passage des composantes *RVB* d'une caméra quelconque aux composantes CIE 1964 *XYZ* peut être modélisé par l'application d'une matrice  $3 \times 3$  (équation (2.11)). Si les sensibilités des capteurs de la caméra connus, cette transformation peut être affinée.
- Réflexion spéculaire neutre (équation (2.7)) :  $F_{spec}(\mathbf{x}, \lambda) = F_{spec}(\mathbf{x})$ .

Geusebroek *et al.* proposent ensuite d'utiliser le modèle couleur gaussien pour définir des caractéristiques couleur invariantes. Nous considérons le stimulus couleur  $C(\mathbf{x}, \lambda)$  réfléchi par la surface élémentaire  $\mathbf{x}$  observée sur le pixel  $P$ . Le

modèle couleur gaussien considère trois capteurs  $G$ ,  $G_\lambda$  et  $G_{\lambda\lambda}$  dont les sensibilités spectrales correspondent respectivement aux dérivées d'ordre 0, 1 et 2 d'une fonction gaussienne  $G(\lambda)$  de longueur d'onde centrale  $\lambda_0 = 520 \text{ nm}$  et d'écart-type  $\sigma_\lambda = 55 \text{ nm}$  [44]. Geusebroek *et al.* ont montré que les composantes couleur exprimées dans cet espace gaussien représentent les trois premiers coefficients du développement de Taylor du stimulus couleur  $C(\mathbf{x}, \lambda)$  pondéré par la gaussienne  $G(\lambda)$ . Autrement dit, ces composantes représentent les trois premières dérivées spectrales du stimulus couleur considéré. Les composantes couleur  $\{C^G(P), C^{G_\lambda}(P), C^{G_{\lambda\lambda}}(P)\}$  du pixel  $P$  dans cet espace gaussien peuvent être approchées à partir des composantes couleur de ce pixel exprimées dans l'espace CIE 1964 XYZ par la transformation suivante :

$$\begin{bmatrix} C^G(P) \\ C^{G_\lambda}(P) \\ C^{G_{\lambda\lambda}}(P) \end{bmatrix} = \begin{pmatrix} -0.48 & 1.2 & 0.28 \\ 0.48 & 0 & -0.4 \\ 1.18 & -1.3 & 0 \end{pmatrix} \begin{bmatrix} C^X(P) \\ C^Y(P) \\ C^Z(P) \end{bmatrix}. \quad (2.49)$$

Ainsi, d'après l'hypothèse d'estimation des composantes XYZ à partir des composantes RVB de la caméra (équation (2.11)), la transformation globale des composantes définies dans l'espace  $RVB$  des capteurs d'acquisition aux composantes définies dans l'espace couleur gaussien est la suivante :

$$\begin{bmatrix} C^G(P) \\ C^{G_\lambda}(P) \\ C^{G_{\lambda\lambda}}(P) \end{bmatrix} = \begin{pmatrix} 0.06 & 0.63 & 0.27 \\ 0.3 & 0.04 & -0.35 \\ 0.34 & -0.6 & 0.17 \end{pmatrix} \begin{bmatrix} C^R(P) \\ C^V(P) \\ C^B(P) \end{bmatrix}. \quad (2.50)$$

Partant de ces résultats, Geusebroek *et al.* définissent de nouvelles caractéristiques couleur invariantes fondées sur les dérivées spectrales du stimulus couleur  $C_{Kub-Munk}(\mathbf{x}, \lambda)$  fournit par le modèle de Kubelka-Munk.

- *Caractéristique couleur invariante H* [45]

Pour la caractéristique invariante  $H$ , Geusebroek *et al.* ajoutent aux 3 hypothèses communes l'hypothèse de l'éclairage blanc (équation (2.17)) :  $E(\mathbf{x}, \lambda) = E(\mathbf{x})$ .

Sous toutes ces conditions, le stimulus couleur  $C_{Kub-Munk}(\mathbf{x}, \lambda)$  réfléchi par la surface élémentaire  $\mathbf{x}$  est défini par l'équation :

$$C_{Kub-Munk}(\mathbf{x}, \lambda) = (1 - F_{spec}(\mathbf{x}))^2 \cdot \beta(\mathbf{x}, \lambda) \cdot E(\mathbf{x}) + F_{spec}(\mathbf{x}) \cdot E(\mathbf{x}) \quad (2.51)$$

Ainsi, les dérivées spectrales première et seconde de ce signal nous donnent :

$$C_{Kub-Munk_\lambda}(\mathbf{x}, \lambda) = (1 - F_{spec}(\mathbf{x}))^2 \cdot \frac{\partial \beta(\mathbf{x}, \lambda)}{\partial \lambda} \cdot E(\mathbf{x}) \quad (2.52)$$

et

$$C_{Kub-Munk_{\lambda\lambda}}(\mathbf{x}, \lambda) = (1 - F_{spec}(\mathbf{x}))^2 \cdot \frac{\partial^2 \beta(\mathbf{x}, \lambda)}{\partial \lambda^2} \cdot E(\mathbf{x}) \quad (2.53)$$

Ainsi, le rapport entre la dérivée première et la dérivée seconde du stimulus couleur ne dépend que de la réflectance spectrale de la surface élémentaire observée. Le modèle couleur gaussien permet d'obtenir ces dérivées et Geusebroek *et al.* proposent donc la caractéristique couleur invariante suivante [45] :

$$X^H(P) = \frac{C^{G_\lambda}(P)}{C^{G_{\lambda\lambda}}(P)} \quad (2.54)$$

Ainsi, lorsque la réflexion spéculaire est neutre et que l'éclairage est blanc, Geusebroek *et al.* ont montré que, pour un même pixel  $P$ , le rapport entre sa composante  $C^{G_\lambda}(P)$  et sa composante  $C^{G_{\lambda\lambda}}(P)$  ne dépend pas de l'intensité de l'éclairage, de la direction de l'éclairage, de l'angle d'observation et des reflets éventuels.

- *Caractéristique couleur invariante C [45]*

Pour la caractéristique invariante  $C$ , Geusebroek *et al.* ajoutent aux 3 hypo-

thèses communes les 2 hypothèses suivantes :

- Surface matte (équation (2.8)) :  $F_{spec}(\mathbf{x}, \lambda) = 0$ . Dans ce cas, l'hypothèse de réflexion spéculaire neutre n'est plus nécessaire.
- Éclairage blanc (équation (2.17)) :  $E(\mathbf{x}, \lambda) = E(\mathbf{x})$

Sous toutes ces conditions, le stimulus couleur  $C_{Kub-Munk}(\mathbf{x}, \lambda)$  réfléchi par la surface élémentaire  $\mathbf{x}$  est défini par l'équation :

$$C_{Kub-Munk}(\mathbf{x}, \lambda) = \beta(\mathbf{x}, \lambda).E(\mathbf{x}). \quad (2.55)$$

La dérivée spectrale première de ce signal donne alors :

$$C_{Kub-Munk_\lambda}(\mathbf{x}, \lambda) = \frac{\partial \beta(\mathbf{x}, \lambda)}{\partial \lambda}.E(\mathbf{x}). \quad (2.56)$$

Le rapport entre la dérivée spectrale première du stimulus couleur et le stimulus lui même ne dépend donc que de la réflectance de la surface élémentaire observée. Geusebroek *et al.* proposent donc la caractéristique couleur invariante suivante :

$$X^C(P) = \frac{C^{G_\lambda}(P)}{C^G(P)}. \quad (2.57)$$

Ainsi, lorsque la surface est matte et que l'éclairage est blanc, Geusebroek *et al.* ont montré que, pour un même pixel  $P$ , le rapport entre sa composante  $C^{G_\lambda}(P)$  et sa composante  $C^G(P)$  ne dépend pas de l'intensité de l'éclairage, de la direction de l'éclairage et de l'angle d'observation.

- *Caractéristique couleur invariante W [45]*

La caractéristique invariante  $W$  fait exception dans ce paragraphe puisqu'elle n'est pas le résultat d'une dérivée spectrale mais d'une dérivée spatiale. Cependant, comme elle est fondée sur les composantes de l'espace gaussien, elle est présentée dans ce paragraphe. Pour cette caractéristique invariante,



Geusebroek *et al.* ajoutent aux 3 hypothèses communes les 3 hypothèses suivantes :

- Surface matte (équation (2.8)) :  $F_{spec}(\mathbf{x}, \lambda) = 0$ . Dans ce cas, l'hypothèse de réflexion spéculaire neutre n'est plus nécessaire.
- Éclairage blanc (équation (2.17)) :  $E(\mathbf{x}, \lambda) = E(\mathbf{x})$ ,
- Éclairage constant localement (équation (2.16)) :  $E(\mathbf{x}, \lambda) = E(\lambda)$ . Lorsque cette hypothèse et la précédente sont vraies, nous avons :  $E(\mathbf{x}, \lambda) = E$ .

Sous toutes ces conditions, le stimulus couleur  $C_{Kub-Munk}(\mathbf{x}, \lambda)$  réfléchi par la surface élémentaire  $\mathbf{x}$  est défini par l'équation :

$$C_{Kub-Munk}(\mathbf{x}, \lambda) = \beta(\mathbf{x}, \lambda).E. \quad (2.58)$$

La dérivée spatiale première de ce signal donne alors :

$$C_{Kub-Munk_{\mathbf{x}}}(\mathbf{x}, \lambda) = \frac{\partial \beta(\mathbf{x}, \lambda)}{\partial \mathbf{x}}.E. \quad (2.59)$$

Le rapport entre la dérivée spatiale première du stimulus couleur et le stimulus lui-même ne dépend donc que de la réflectance de la surface élémentaire observée. Geusebroek *et al.* proposent donc la caractéristique couleur invariante suivante [45] :

$$X^W(P) = \frac{C^{G_{\mathbf{x}}}(P)}{C^G(P)}, \quad (2.60)$$

où  $C^{G_{\mathbf{x}}}(P) = \frac{\partial C^G(P)}{\partial \mathbf{x}}$  représente la dérivée spatiale de la composante  $C^G(P)$  en  $P$ .

Ainsi, lorsque la surface est matte, que l'éclairage est blanc et constant localement, Geusebroek *et al.* ont montré que, pour un même pixel  $P$ , le rapport entre la dérivée spatiale de la composante  $C^G(P)$  et la composante  $C^G(P)$  elle-même ne dépend pas de l'intensité de l'éclairage.

- *Caractéristique couleur invariante N [45]*

Pour la caractéristique invariante  $N$ , Geusebroek *et al.* ajoutent aux 3 hypothèses communes les 2 hypothèses suivantes :

- Surface matte (équation (2.8)) :  $F_{spec}(\mathbf{x}, \lambda) = 0$ . Dans ce cas, l'hypothèse de réflexion spéculaire neutre n'est plus nécessaire.
- Couleur de l'éclairage constante localement (équation (2.15)) :  $E(\mathbf{x}, \lambda) = e(\mathbf{x}).E(\lambda)$ .

Sous toutes ces conditions, le stimulus couleur  $C_{Kub-Munk}(\mathbf{x}, \lambda)$  réfléchi par la surface élémentaire  $\mathbf{x}$  est défini par l'équation :

$$C_{Kub-Munk}(\mathbf{x}, \lambda) = \beta(\mathbf{x}, \lambda).e(\mathbf{x}).E(\lambda). \quad (2.61)$$

La dérivée spectrale première de ce signal donne alors :

$$C_{Kub-Munk_\lambda}(\mathbf{x}, \lambda) = e(\mathbf{x}) \left[ \frac{\partial \beta(\mathbf{x}, \lambda)}{\partial \lambda} . E(\lambda) + \frac{\partial E(\lambda)}{\partial \lambda} . \beta(\mathbf{x}, \lambda) \right]. \quad (2.62)$$

Le rapport entre la dérivée spectrale première du stimulus couleur et le stimulus lui même est égal à :

$$\frac{C_{Kub-Munk_\lambda}(\mathbf{x}, \lambda)}{C_{Kub-Munk}(\mathbf{x}, \lambda)} = \frac{1}{\beta(\mathbf{x}, \lambda)} \frac{\partial \beta(\mathbf{x}, \lambda)}{\partial \lambda} + \frac{1}{E(\lambda)} \frac{\partial E(\lambda)}{\partial \lambda}. \quad (2.63)$$

Par conséquent, si nous dérivons l'équation précédente par rapport à  $\mathbf{x}$ , la partie qui dépend de l'éclairage va disparaître et nous obtiendrons une valeur qui ne dépend que de la réflectance spectrale de la surface élémentaire :

$$\frac{\partial}{\partial \mathbf{x}} \frac{C_{Kub-Munk_\lambda}(\mathbf{x}, \lambda)}{C_{Kub-Munk}(\mathbf{x}, \lambda)} = \frac{\partial}{\partial \mathbf{x}} \frac{1}{\beta(\mathbf{x}, \lambda)} \frac{\partial \beta(\mathbf{x}, \lambda)}{\partial \lambda}. \quad (2.64)$$

Geusebroek *et al.* proposent donc la caractéristique invariante suivante :

$$X^N(P) = \frac{C^{G_{\lambda x}}(P)C^G(P) - C^{G_{\lambda}}(P)C^{G_x}(P)}{(C^G(P))^2}, \quad (2.65)$$

où  $C^{G_x}(P) = \frac{\partial C^G(P)}{\partial \mathbf{x}}$  et  $C^{G_{\lambda x}}(P) = \frac{\partial C^{G_{\lambda}}(P)}{\partial \mathbf{x}}$ .

Ainsi, lorsque la surface est matte et que la couleur de l'éclairage est constante localement, Geusebroek *et al.* ont montré que, pour un même pixel  $P$ , la dérivée spatiale du rapport entre sa composante  $C^{G_{\lambda}}(P)$  et sa composante  $C^G(P)$  ne dépend pas de l'intensité, de la direction ou de la couleur de l'éclairage et de l'angle d'observation.

#### (b) Les dérivées spatiales de van de Weijer *et al.* [99, 100]

Van de Weijer *et al.* proposent des caractéristiques invariantes qui reposent sur les quatre hypothèses suivantes :

- Modèle dichromatique de Shafer (équation (2.4)) :  

$$C_{Shafer}(\mathbf{x}, \lambda) = m_{diff}(\theta) \cdot \beta(\mathbf{x}, \lambda) \cdot E(\mathbf{x}, \lambda) + m_{spec}(\theta, \alpha) \cdot F_{spec}(\mathbf{x}, \lambda) \cdot E(\mathbf{x}, \lambda).$$
- Réflexion spéculaire neutre (équation (2.7)) :  $F_{spec}(\mathbf{x}, \lambda) = F_{spec}(\mathbf{x})$ .
- Couleur de l'éclairage constante localement (équation (2.15)) :  

$$E(\mathbf{x}, \lambda) = e(\mathbf{x}) \cdot E(\lambda).$$
- Pour déterminer les caractéristiques invariantes, ils supposent que la couleur de l'illuminant est connue (hypothèse 13 page 31), mais pas l'intensité de l'éclairage. Si la couleur n'est pas connue, ils supposent que l'éclairage est blanc (équation (2.17)) :  $E(\mathbf{x}, \lambda) = E(\mathbf{x})$ . Pour la suite de la démonstration, nous allons partir de l'hypothèse d'un spectre connu non plat.

Ainsi, les composantes couleur  $C^k(P)$ ,  $k = R, V, B$ , du pixel  $P$  qui observe la surface élémentaire à la position  $\mathbf{x}$  sont estimées par :

$$\begin{aligned} C^k(P) &= m_{diff}(\theta)e(\mathbf{x}) \int_{\lambda} \beta(\mathbf{x}, \lambda).E(\lambda)d\lambda + \\ &\quad m_{spec}(\theta, \alpha).F_{spec}(\mathbf{x}).e(\mathbf{x}) \int_{\lambda} E(\lambda)d\lambda \quad (2.66) \\ &= e(\mathbf{x})(m_{diff}(\theta)C_{diff}^k(P) + n_{spec}(\theta, \alpha, \mathbf{x})C_{spec}^k), \end{aligned}$$

où

$$\begin{cases} C_{diff}^k(P) = \int_{\lambda} \beta(\mathbf{x}, \lambda).E(\lambda).k(\lambda)d\lambda, \\ C_{spec}^k = \int_{\lambda} E(\lambda).k(\lambda)d\lambda, \\ n_{spec}(\theta, \alpha, \mathbf{x}) = m_{spec}(\theta, \alpha).F_{spec}(\mathbf{x}). \end{cases} \quad (2.67)$$

Si nous dérivons l'expression précédente par rapport à  $\mathbf{x}$ , nous obtenons :

$$\begin{aligned} C_{\mathbf{x}}^k(P) &= e_{\mathbf{x}}(\mathbf{x})(m_{diff}(\theta)C_{diff}^k(P) + n_{spec}(\theta, \alpha, \mathbf{x})C_{spec}^k) \\ &\quad + e(\mathbf{x})(m_{diff_{\mathbf{x}}}(\theta)C_{diff}^k(P) + m_{diff}(\theta)C_{diff_{\mathbf{x}}}^k(P) \\ &\quad + n_{spec_{\mathbf{x}}}(\theta, \alpha, \mathbf{x})C_{spec}^k), \quad (2.68) \\ &= (e(\mathbf{x})m_{diff}(\theta))C_{diff_{\mathbf{x}}}^k(P) \\ &\quad + (e_{\mathbf{x}}(\mathbf{x})m_{diff}(\theta) + e(\mathbf{x})m_{diff_{\mathbf{x}}}(\theta))C_{diff}^k(P) \\ &\quad + (e_{\mathbf{x}}(\mathbf{x})n_{spec}(\theta, \alpha, \mathbf{x}) + e(\mathbf{x})n_{spec_{\mathbf{x}}}(\theta, \alpha, \mathbf{x}))C_{spec}^k, \end{aligned}$$

où l'indice  $\mathbf{x}$  définit la dérivée spatiale.

Par conséquent, la dérivée spatiale  $\mathbf{C}_{\mathbf{x}}(P) = (C_{\mathbf{x}}^R(P), C_{\mathbf{x}}^V(P), C_{\mathbf{x}}^B(P))^t$  du vecteur couleur  $\mathbf{C}(P) = (C^R(P), C^V(P), C^B(P))^t$ , du pixel  $P$  qui observe la surface élémentaire à la position  $\mathbf{x}$  est exprimée par :

$$\begin{aligned} \mathbf{C}_{\mathbf{x}}(P) &= (e(\mathbf{x})m_{diff}(\theta))\mathbf{C}_{diff_{\mathbf{x}}}(P) \\ &\quad + (e_{\mathbf{x}}(\mathbf{x})m_{diff}(\theta) + e(\mathbf{x})m_{diff_{\mathbf{x}}}(\theta))\mathbf{C}_{diff}(P) \quad (2.69) \\ &\quad + (e_{\mathbf{x}}(\mathbf{x})n_{spec}(\theta, \alpha, \mathbf{x}) + e(\mathbf{x})n_{spec_{\mathbf{x}}}(\theta, \alpha, \mathbf{x}))\mathbf{C}_{spec}. \end{aligned}$$

Cela signifie que la dérivée spatiale du vecteur couleur dans une image est la somme pondérée de 3 vecteurs. Van de Weijer *et al.* associent un phénomène physique à chacune des trois directions correspondantes :

- $(e(\mathbf{x})m_{diff}(\theta))\mathbf{C}_{diff_x}(P)$  est associée à la différence de reflectances spectrales des surfaces considérées pour la dérivée.
- $(e_x(\mathbf{x})m_{diff}(\theta) + e(\mathbf{x})m_{diff_x}(\theta))\mathbf{C}_{diff}(P)$  est associée à la différence d'ombres sur les surfaces considérées. L'ombre peut être expliquée par la présence d'un objet entre l'éclairage et la surface considérée, Van de Weijer *et al.* associent ce phénomène au scalaire  $e_x(\mathbf{x})m_{diff}(\theta)$ . Mais elle peut être aussi expliquée par une variation de  $\theta$ , l'angle formé par le rayon lumineux et la normale à la surface, Van de Weijer *et al.* associent ce phénomène au scalaire  $e(\mathbf{x})m_{diff_x}(\theta)$ . Nous pouvons noter que s'il n'y a pas de réflexion spéculaire sur les surfaces considérées, le vecteur couleur du pixel  $P$  est alors  $\mathbf{C}(P) = e(\mathbf{x})(m_{diff}(\theta)\mathbf{C}_{diff}(P))$ . Par conséquent, le vecteur  $(e_x(\mathbf{x})m_{diff}(\theta) + e(\mathbf{x})m_{diff_x}(\theta))\mathbf{C}_{diff}(P)$  expliqué par les différences d'ombres entre les surfaces a la même direction que le vecteur couleur de la surface considérée si celle-ci est matte (réflexion spéculaire nulle).
- $(e_x(\mathbf{x})n_{spec}(\theta, \alpha, \mathbf{x}) + e(\mathbf{x})n_{spec_x}(\theta, \alpha, \mathbf{x}))\mathbf{C}_{spec}$  est associée à la différence de réflexion spéculaire entre les surfaces considérées. De la même manière, cette variation peut être expliquée par deux phénomènes. Le scalaire  $e_x n_{spec}(\theta, \alpha, \mathbf{x})$  correspond à un contour d'ombre superposé à une réflexion spéculaire alors que le scalaire  $e(\mathbf{x})n_{spec_x}(\theta, \alpha, \mathbf{x})$  correspond à une variation d'un ou plusieurs des angles suivants : direction de l'éclairage, point de vue ou orientation de la surface. Comme  $C_{spec}^k = \int_{\lambda} E(\lambda) \cdot k(\lambda) d\lambda$ , nous en déduisons que la direction

de ce vecteur associé à la différence de réflexion spéculaire est celle de la couleur de l'illuminant.

Par cette analyse, van de Weijer *et al.* montrent que, à partir de la couleur du pixel  $\mathbf{C}(P) = (C^R(P), C^V(P), C^B(P))^t$  et de la couleur de l'éclairage  $(\int_{\lambda} E(\lambda) \cdot R(\lambda) d\lambda, \int_{\lambda} E(\lambda) \cdot V(\lambda) d\lambda, \int_{\lambda} E(\lambda) \cdot B(\lambda) d\lambda)^t = (E^R, E^V, E^B)^t$ , il est possible de calculer la direction (dans l'espace couleur) de deux des causes d'un contour :

- direction de l'ombre :  $\mathbf{O}(P) = \frac{(C^R(P), C^V(P), C^B(P))^t}{\sqrt{C^R(P)^2 + C^V(P)^2 + C^B(P)^2}}$ ,
- direction spéculaire :  $\mathbf{S} = \frac{(E^R, E^V, E^B)^t}{\sqrt{(E^R)^2 + (E^V)^2 + (E^B)^2}}$ .

A partir de ces deux directions, il est possible de créer une troisième direction  $\mathbf{T}$  qui est perpendiculaire à ces deux vecteurs et que van de Weijer *et al.* appellent la direction de la teinte :  $\mathbf{T} = \frac{\mathbf{O} \times \mathbf{S}}{|\mathbf{O} \times \mathbf{S}|}$ , où  $\times$  représente le produit vectoriel. Van de Weijer *et al.* notent que cette direction n'est pas forcément celle associée à la différence de réflectance spectrale décrite plus haut mais que de l'analyse précédente, ils déduisent également que les variations dans la direction de la teinte  $\mathbf{T}$  peuvent être attribués uniquement à une modification de la réflectance spectrale des surfaces considérées. Klinker *et al.* [61] avaient déjà défini ces directions pour la segmentation d'images.

- *Invariance à l'ombre [99]*

Si les contours associés à une variation de l'ombre ont une direction (dans l'espace couleur) identique à la direction de l'ombre  $\mathbf{O}$ , alors lorsque l'on calcule la dérivée couleur  $\mathbf{C}_x(P)$  en un pixel  $P$ , la projection du vecteur  $\mathbf{C}_x(P)$  sur le vecteur  $\mathbf{O}(P)$  :  $(\mathbf{C}_x(P) \cdot \mathbf{O}(P)) \mathbf{O}(P)$  nous donne la part du contour qui peut être expliquée par une variation de l'ombre. Si l'on soustrait cette projection au vecteur dérivée  $\mathbf{C}_x(P)$ , alors le résultat est indépendant des ombres. Par conséquent, pour déterminer une caractéristique

couleur  $\mathbf{X}^{oq}(P)$  invariante à l'ombre, van de Weijer *et al.* proposent de projeter la dérivée du vecteur couleur sur la direction de l'ombre  $\mathbf{O}(P)$  et de soustraire ce résultat à cette dérivée :

$$\mathbf{X}^{oq}(P) = \mathbf{C}_x(P) - (\mathbf{C}_x(P) \cdot \mathbf{O}(P))\mathbf{O}(P). \quad (2.70)$$

Cette caractéristique représente la partie de la dérivée qui n'est pas provoquée par un contour d'ombre, elle représente donc uniquement les contours de reflectance spectrale et les contours de réflexion spéculaire.

Van de Weijer *et al.* montrent que cette caractéristique peut être obtenue par dérivation spatiale des composantes de l'espace couleur sphérique  $r\theta\phi$  [99].

De plus, de manière à rendre cette caractéristique invariante à la direction et à l'intensité de l'éclairage, ils proposent de la diviser par la norme du vecteur couleur du pixel considéré [100] :

$$\mathbf{X}^{of}(P) = \frac{\mathbf{X}^{oq}(P)}{|\mathbf{C}(P)|}. \quad (2.71)$$

La caractéristique  $\mathbf{X}^{of}(P)$  est appelée "shadow-shading full invariant" alors que  $\mathbf{X}^{oq}(P)$  est appelée "shadow-shading quasi-invariant" [100]. Les auteurs montrent que les "quasi-invariants" sont plus stables que les "full invariants" en cas de bruit et sont d'avantage dédiés à l'extraction de points caractéristiques qu'à la description des régions locales autour de ces points.

- *Invariance aux reflets [99]*

Pour déterminer une caractéristique couleur  $\mathbf{X}^{sq}(P)$  invariante aux reflets, van de Weijer *et al.* proposent de projeter la dérivée du vecteur couleur sur la direction spéculaire  $\mathbf{S}$  et de soustraire ce résultat à cette dérivée :

$$\mathbf{X}^{sq}(P) = \mathbf{C}_x(P) - (\mathbf{C}_x(P) \cdot \mathbf{S}(P))\mathbf{S}(P). \quad (2.72)$$

Cette caractéristique représente la partie de la dérivée qui n'est pas provoquée par un contour de réflexion spéculaire, elle représente donc uniquement les contours de reflectance spectrale et d'ombre.

Van de Weijer *et al.* montrent que cette caractéristique peut être obtenue par dérivation spatiale des composantes de l'espace couleur opposé  $o_1 o_2 o_3$  [99].

- *Invariance à l'ombre et aux reflets [99]*

Pour déterminer une caractéristique couleur  $\mathbf{X}^{soq}(P)$  invariante à l'ombre et aux reflets, van de Weijer *et al.* proposent de projeter la dérivée du vecteur couleur sur la direction de la teinte  $\mathbf{T}(P)$  [99] :

$$\mathbf{X}^{soq}(P) = (\mathbf{C}_x(P) \cdot \mathbf{T}(P)) \mathbf{T}(P). \quad (2.73)$$

Cette caractéristique représente la partie de la dérivée qui n'est provoquée ni par un contour d'ombre, ni par un contour de réflexion spéculaire, elle représente donc uniquement les contours de reflectance spectrale.

Van de Weijer *et al.* montrent que cette caractéristique peut être obtenue par dérivation spatiale des composantes de l'espace couleur *hsi* [99].

De plus, de manière à rendre cette caractéristique invariante à la direction et à l'intensité de l'éclairage et au point de vue de l'observateur, ils proposent de la diviser par la saturation de la couleur du pixel considéré. La saturation est alors définie comme la norme du vecteur couleur  $\mathbf{C}(P)$  après projection sur le plan perpendiculaire à la direction spéculaire  $\mathbf{S}$  :

$$\mathbf{X}^{sof}(P) = \frac{\mathbf{X}^{soq}(P)}{|\mathbf{C}(P) - (\mathbf{C}(P) \cdot \mathbf{S}) \mathbf{S}|}. \quad (2.74)$$

La caractéristique  $\mathbf{X}^{sof}(P)$  est appelée "shadow-shading-specular full invariant" alors que  $\mathbf{X}^{soq}(P)$  est appelée "shadow-shading-specular quasi-invariant".



Les sections 2.3.1 à 2.3.3 ont été consacrées à lister et décrire une grande partie des caractéristiques couleur invariantes utilisées dans les applications de reconnaissance d'objets. Nous proposons maintenant, dans la section suivante, de résumer, pour chaque caractéristique, les hypothèses de base ainsi que les variations photométriques auxquelles elles sont invariantes.

### 2.3.4 Synthèse sur les hypothèses et les invariances des caractéristiques couleur

Le tableau 2.1 reprend toutes les caractéristiques invariantes présentées dans les sections précédentes. A chacune est associée une dénomination du type  $Car_i$ , où  $i$  est un nombre entre 1 et 21. Le tableau 2.2 reprend toutes les hypothèses et leur associe une dénomination du type  $H_j$  où  $j$  est un nombre entre 1 et 18. Ces dénominations sont alors utilisées dans les tableaux 2.3 et 2.4 afin de synthétiser respectivement les hypothèses de base sur lesquelles sont fondées chaque caractéristique et les variations photométriques et radiométriques auxquelles elles sont invariantes. Ainsi, en fonction des particularités d'une application, notamment les conditions d'acquisition des images et des invariances recherchées, ces tableaux vont permettre de sélectionner les caractéristiques couleurs les plus appropriées. Il faut préciser que le tableau 2.4 ne peut pas être utilisé sans le tableau 2.3, car certaines invariances sont obtenues uniquement parce que certaines hypothèses sont vérifiées. Par exemple, lorsque l'hypothèse de surface lambertienne est vérifiée, alors les caractéristiques que l'on peut retenir sont de fait insensibles aux changements de point de vue. Cette invariance n'est pas obtenue par une normalisation de la caractéristique couleur utilisée mais par la validation de l'hypothèse. Par conséquent, il faut vérifier que les hypothèses d'application (tableau 2.3) d'une caractéristique couleur sont bien validées pour être sûr que l'invariance prévue dans le tableau 2.4 soit effective. Rappelons enfin que dans cette étude, nous n'avons pas présenté l'invariance à l'orientation de la surface comme dans certains articles car

## CHAPITRE 2. LES DESCRIPTEURS COULEUR INVARIANTS DE LA LITTÉRATURE

---

celle-ci est redondante avec les invariances au point de vue et à la direction de l'éclairage. En effet, lorsqu'une caractéristique couleur est invariante au point de vue et à la direction de l'éclairage, elle est de fait aussi invariante à l'orientation de la surface observée.

Dénomination dans les tableaux suivants	Description et/ou Référence	Equation et/ou page
$Car_1$	Les rapports entre composantes de Funt <i>et al.</i> [41]	Equation 2.28 page 37
$Car_2$	Les caractéristiques $m_1, m_2, m_3$ de Gevers <i>et al.</i> [48]	Equation 2.34 page 41
$Car_3$	Les caractéristiques $l_1, l_2, l_3$ de Gevers <i>et al.</i> [48]	Equation 2.37 page 42
$Car_4$	Les caractéristiques $c_1, c_2, c_3$ de Gevers <i>et al.</i> [48]	Equation 2.41 page 43
$Car_5$	La minimisation de l'entropie : Finlayson <i>et al.</i> [34, 35]	Page 46
$Car_6$	L'égalisation d'histogramme de Finlayson <i>et al.</i> [28]	Page 47
$Car_7$	La normalisation des moments de Lenz <i>et al.</i> [65]	Page 48
$Car_8$	La normalisation des valeurs propres de Healey <i>et al.</i> [49]	Page 48
$Car_9$	La normalisation de la moyenne et de l'écart-type	Equation 2.47 page 49
$Car_{10}$	Les moments couleur invariants au modèle diagonal de Mindru <i>et al.</i> [77]	Page 50
$Car_{11}$	Les moments couleur invariants au modèle diagonal et translation de Mindru <i>et al.</i> [77]	Page 50
$Car_{12}$	Les moments couleur invariants au modèle affine et translation de Mindru <i>et al.</i> [77]	Page 50
$Car_{13}$	La caractéristique $H$ de Geusebroek <i>et al.</i> [45]	Equation 2.54 page 52
$Car_{14}$	La caractéristique $C$ de Geusebroek <i>et al.</i> [45]	Equation 2.57 page 53
$Car_{15}$	La caractéristique $W$ de Geusebroek <i>et al.</i> [45]	Equation 2.60 page 54
$Car_{16}$	La caractéristique $N$ de Geusebroek <i>et al.</i> [45]	Equation 2.65 page 56
$Car_{17}$	La caractéristique "quasi-invariant" aux ombres de van de Weijer <i>et al.</i> [99]	Equation 2.70 page 60
$Car_{18}$	La caractéristique "full-invariant" aux ombres de van de Weijer <i>et al.</i> [99]	Equation 2.71 page 60
$Car_{19}$	La caractéristique "quasi-invariant" aux reflets de van de Weijer <i>et al.</i> [99]	Equation 2.72 page 60
$Car_{20}$	La caractéristique "quasi-invariant" aux ombres et aux reflets de van de Weijer <i>et al.</i> [99]	Equation 2.73 page 61
$Car_{21}$	La caractéristique "full-invariant" aux ombres et aux reflets de van de Weijer <i>et al.</i> [99]	Equation 2.74 page 61

TABLE 2.1 – Caractéristiques couleur invariantes présentées dans le paragraphe précédent

## CHAPITRE 2. LES DESCRIPTEURS COULEUR INVARIANTS DE LA LITTÉRATURE

---

Dénomination de l'hypothèse	Description de l'hypothèse et page
$H_1$	Modèle de Kubelka-Munk, page 25
$H_2$	Modèle de Shafer, page 25
$H_3$	Modèle Lambertien, page 26
$H_4$	Réflexion spéculaire neutre, page 27
$H_5$	Surface Matte, page 27
$H_6$	Intégrales des sensibilités des 3 capteurs identiques, page 27
$H_7$	Bandes-passantes des capteurs étroites, page 28
$H_8$	Transformation de l'espace RVB de la caméra à l'espace CIE XYZ, page 28
$H_9$	Equation de Planck, page 29
$H_{10}$	Couleur de l'éclairage constante localement, page 29
$H_{11}$	Éclairage uniforme localement, page 30
$H_{12}$	Illuminant blanc, page 31
$H_{13}$	Couleur de l'éclairage connue, page 31
$H_{14}$	Modèle diagonale, page 32
$H_{15}$	Modèle diagonale et translation, page 33
$H_{16}$	Transformation affine, page 33
$H_{17}$	Transformation affine et translation, page 34
$H_{18}$	Fonctions monotones croissantes, page 35

TABLE 2.2 – Hypothèses de base utilisées par les caractéristiques couleur

Caractéristique	$H_1$	$H_2$	$H_3$	$H_4$	$H_5$	$H_6$	$H_7$	$H_8$	$H_9$	$H_{10}$	$H_{11}$	$H_{12}$	$H_{13}$	$H_{14}$	$H_{15}$	$H_{16}$	$H_{17}$	$H_{18}$
$Car_1$			$\oplus$				$\oplus$				$\oplus$							
$Car_2$		$\oplus$			$\oplus$		$\oplus$			$\oplus$								
$Car_3$		$\oplus$		$\oplus$		$\oplus$						$\oplus$						
$Car_4$		$\oplus$			$\oplus$							$\oplus$						
$Car_5$			$\oplus$				$\oplus$		$\oplus$									
$Car_6$																		$\oplus$
$Car_7$																$\oplus$		
$Car_8$																$\oplus$		
$Car_9$															$\oplus$			
$Car_{10}$														$\oplus$				
$Car_{11}$															$\oplus$			
$Car_{12}$																	$\oplus$	
$Car_{13}$	$\oplus$			$\oplus$				$\oplus$				$\oplus$						
$Car_{14}$	$\oplus$				$\oplus$			$\oplus$				$\oplus$						
$Car_{15}$	$\oplus$				$\oplus$						$\oplus$	$\oplus$						
$Car_{16}$	$\oplus$				$\oplus$			$\oplus$		$\oplus$								
$Car_{17}$		$\oplus$		$\oplus$						$\oplus$			$\oplus$					
$Car_{18}$		$\oplus$		$\oplus$						$\oplus$			$\oplus$					
$Car_{19}$		$\oplus$		$\oplus$						$\oplus$			$\oplus$					
$Car_{20}$		$\oplus$		$\oplus$						$\oplus$			$\oplus$					
$Car_{21}$		$\oplus$		$\oplus$						$\oplus$			$\oplus$					

TABLE 2.3 – Hypothèses sur lesquelles reposent chaque caractéristique couleur invariante

Caractéristique	Intensité de l'éclairage	Couleur de l'éclairage	Direction de l'éclairage	Point de vue	Lumière ambiante	Ombres	Reflets
<i>Car</i> <sub>1</sub>	⊕	⊕		⊕			
<i>Car</i> <sub>2</sub>	⊕	⊕	⊕	⊕			
<i>Car</i> <sub>3</sub>	⊕		⊕	⊕			⊕
<i>Car</i> <sub>4</sub>	⊕		⊕	⊕			
<i>Car</i> <sub>5</sub>	⊕	⊕				⊕	
<i>Car</i> <sub>6</sub>	⊕	⊕					
<i>Car</i> <sub>7</sub>	⊕	⊕					
<i>Car</i> <sub>8</sub>	⊕	⊕					
<i>Car</i> <sub>9</sub>	⊕	⊕			⊕		
<i>Car</i> <sub>10</sub>	⊕	⊕					
<i>Car</i> <sub>11</sub>	⊕	⊕			⊕		
<i>Car</i> <sub>12</sub>	⊕	⊕			⊕		
<i>Car</i> <sub>13</sub>	⊕		⊕	⊕			⊕
<i>Car</i> <sub>14</sub>	⊕		⊕	⊕			
<i>Car</i> <sub>15</sub>	⊕						
<i>Car</i> <sub>16</sub>	⊕	⊕	⊕	⊕			
<i>Car</i> <sub>17</sub>						⊕	
<i>Car</i> <sub>18</sub>	⊕		⊕	⊕		⊕	
<i>Car</i> <sub>19</sub>							⊕
<i>Car</i> <sub>20</sub>						⊕	⊕
<i>Car</i> <sub>21</sub>	⊕		⊕	⊕		⊕	⊕

TABLE 2.4 – Phénomènes photométriques et radiométriques auxquels sont invariantes chaque caractéristique

Après avoir présenté les principales caractéristiques couleur invariantes, nous proposons maintenant de montrer comment la couleur peut être prise en compte dans les descripteurs locaux de couleur.

## 2.4 Les descripteurs locaux couleur

Dans la continuité des sections 2.2 et 2.3, l'objet de cette section est de lister l'ensemble des descripteurs locaux qui peuvent être utilisés dans le cadre de la reconnaissance d'image, et de présenter leur particularité.

Le descripteur SIFT qui est le plus largement utilisé [68] est décrit dans l'Annexe B. Ce descripteur est calculé à partir des d'histogrammes locaux des orientations de gradients. Il exploite donc avant tout l'information de forme. Une étude récente a comparé les taux de reconnaissance d'objets à partir du descripteur SIFT avec ceux obtenus à partir des descripteurs couleur classiques, comme les histogrammes couleur calculés dans différents espaces colorimétriques ou les moments couleur, les résultats ont montré la supériorité du descripteur SIFT [98]. Pourtant, les images sur lesquelles ont été testés ces descripteurs présentes des objets pour lesquels l'information de couleur paraît discriminante. Il semble donc opportun de combiner l'information de couleur et l'information de forme pour augmenter les taux de reconnaissance. Dans cette section, nous proposons de présenter de quelle manière l'information de couleur peut être intégrée à un système de reconnaissance d'objets fondé sur l'information de forme. Nous distinguons ici quatre approches principales pour combiner les informations de couleur et de forme. La première consiste simplement à calculer un descripteur de forme et un descripteur couleur et à les concaténer. La seconde repose sur deux étapes successives de comparaison d'images, l'une qui utilise la couleur et l'autre la forme. La troisième met en œuvre en parallèle un descripteur de forme et un descripteur couleur, le plus souvent de manière indépendante, puis fusionne les résultats obtenus pour comparer les images. La dernière approche repose sur des descripteurs spatia-

colorimétriques qui, de par leur structure, représentent à la fois la couleur et la forme des objets. Nous insisterons dans les paragraphes suivants sur le fait que de nombreux descripteurs couleur exploitent les invariances colorimétriques présentées dans les sections précédentes de ce chapitre. Nous ferons référence à ces invariants en donnant à chaque fois le numéro de page à laquelle le lecteur pourra se référer dans les sections précédentes.

### 2.4.1 Concaténation de deux types de descripteurs

Nous allons commencer par décrire quelques méthodes qui associent simplement les deux familles de descripteurs, par concaténation.

Ainsi, Quelhas *et al.* [90], comme la plupart des autres auteurs, proposent d'utiliser le descripteur SIFT pour représenter l'information locale de forme. Le descripteur SIFT ayant une dimension de 128, Quelhas *et al.* proposent d'appliquer une analyse en composantes principales pour réduire cette dimension à 44. L'information de couleur de la région est quant à elle synthétisée par 6 valeurs qui sont sa moyenne et son écart-type calculés sur chacune des composantes de l'espace  $L^*u^*v^*$ . Les auteurs ont proposé de pondérer les deux descripteurs pour la concaténation afin de donner plus ou moins d'importance à la couleur par rapport à la forme. Après différents tests sur la base de Vogel *et al.* [104] ils ont retenu les poids de 0.8 pour la forme et de 0.2 pour la couleur. Sur cette base, les taux de reconnaissance sont de l'ordre de 63% pour la forme seule, 54% pour la couleur seule et 66.5% pour la combinaison proposée.

Van de Weijer *et al.* [101] ont aussi proposé une concaténation du descripteur SIFT avec divers histogrammes couleur locaux. Ces histogrammes sont calculés à partir de différentes composantes couleur :

- les composantes normalisées  $C^r, C^v$  (2D) (cf. page 44),
- la teinte (1D),



- l'angle de composantes opposées (1D) [99], fondé sur un rapport de dérivées dans l'espace couleur opposé  $o_1 o_2 o_3$  (cf. page 61),
- l'angle sphérique (1D) [99], fondé sur un rapport de dérivées dans l'espace couleur sphérique  $r\theta\phi$  (cf. page 60),
- les composantes normalisées par la méthode de Finlayson [37] (cf. page 44).

De plus, les auteurs proposent de pondérer chaque occurrence de ces composantes en fonction de leur instabilité, afin d'obtenir des histogrammes robustes. Ils conservent les 128 dimensions du SIFT et quantifient les histogrammes couleurs sur 37 ou 121 cellules, selon qu'ils considèrent des données 1D ou 2D, respectivement. Les SIFT sont combinés à un seul histogramme couleur à chaque fois, donnant ainsi 5 combinaisons à tester. Expérimentalement, ils ont retenu les poids de 1 pour la forme et 0.6 pour la couleur. Ils ont testé les performances des descripteurs dans 3 conditions différentes : i) éclairage variable et point de vue fixe ; ii) éclairage fixe et point de vue variable et iii) éclairage et point de vue variables. Dans le premier cas, ils montrent que l'addition d'information couleur au descripteur de forme n'apporte que très peu d'information supplémentaire. Dans le second cas, le descripteur couleur donne de meilleurs résultats que le descripteur de forme et l'ajout d'information de forme à celle de couleur ne permet pas d'augmenter les performances. Dans le troisième cas, les auteurs montrent que le meilleur descripteur est celui qui combine la couleur et la forme.

Dahl *et al.* [23] quant à eux ont pris le parti de résumer l'information couleur d'une région détectée à sa moyenne dans l'espace RVB. Au préalable, ils recommandent d'appliquer une égalisation des histogrammes de chaque composante pour obtenir des valeurs peu sensibles aux variations d'illumination [28] (cf. page 47 pour la justification). Le descripteur proposé est une concaténation des SIFT dont la dimension a été réduite à 12 et du descripteur couleur 3D avec les pondérations suivantes : 0.5 pour la forme et 0.5 pour la couleur. Les tests ont été effectués

sur la base de H. Stewénus et D. Nistér (<http://www.vis.uky.edu/stewe/ukbench/>) et montrent que l'information de couleur, même si elle est très simple, permet d'augmenter de 7% les meilleurs résultats obtenus par les SIFT seuls.

De même que Van de Weijer *et al.*, Zhang *et al.* ont combiné les descripteurs SIFT avec des histogrammes couleur locaux [109]. Ils ont choisi de travailler avec les composantes couleur suivantes :

- les composantes normalisées  $C^r, C^v$  quantifiées chacune sur 9 niveaux,
- la teinte quantifiée sur 36 niveaux,
- les composantes  $l_1, l_2$  [48] quantifiées chacune sur 9 niveaux (cf. page 41).

Pour ajouter de l'information spatiale dans la description colorimétrique, les auteurs ont proposé de diviser circulairement la région avoisinante autour des points caractéristiques détectés en 3 anneaux concentriques et de calculer les histogrammes de chacun de ces anneaux. Le descripteur couleur est donc ici une concaténation des 3 histogrammes couleur, donnant ainsi un descripteur à 594 dimensions ( $3 \times 9 \times 9 + 3 \times 36 + 3 \times 9 \times 9$ ). Une ACP permet de réduire cette dimension à 60. Le descripteur forme-couleur est une concaténation du descripteur SIFT avec ce descripteur couleur, donnant ainsi un descripteur de dimension 188. La pondération utilisée ici est de 1 pour la forme et 0.2 pour la couleur. Les tests expérimentaux montrent que sur la base de l'INRIA (<http://lear.inrialpes.fr/data>), l'ajout de la couleur au descripteur SIFT permet d'augmenter le nombre d'appariements corrects entre deux images de 45%.

### 2.4.2 Deux étapes successives pour la comparaison d'images

Passons maintenant aux approches que l'on pourrait qualifier de séquentielles. Les trois méthodes suivantes utilisent la couleur dans une première étape pour comparer les caractéristiques colorimétriques de deux régions image. Des descripteurs de forme sont ensuite calculés, dans une seconde étape, pour vérifier que

ces régions présentent aussi des caractéristiques de forme similaires.

La première méthode, proposée par Khan *et al.* [60, 103] repose sur une analyse locale des distributions des couleurs. Plus précisément, les auteurs proposent de caractériser une région par un histogramme de teintes ou par un histogramme de couleurs spécifiques obtenues par apprentissage [102]. Pour leur tâche de classification d'objets, ils utilisent une approche sac de mots (cf. Annexe A) en calculant des mots visuels liés à la forme ainsi que des mots visuels liés à la couleur. L'idée ici est de pondérer chaque mot visuel de forme dans le sac de mots de forme par la probabilité, fournie par les mots visuels couleur, que le pixel d'où est extrait ce mot appartienne à l'objet recherché. L'intérêt de cette approche repose sur le fait que, si la couleur n'est pas discriminante dans une base d'image, les probabilités couleur seront uniformément réparties sur tous les objets et seule la forme sera prise en compte. Les problèmes liés à la définition des poids optimaux forme/couleur pour chaque base peuvent ainsi être surmontés. Dans le même contexte, Elsayad *et al.* [26] et Chenet *et al.* [20] ont proposé d'utiliser des sacs de mots visuels de forme et de pondérer la contribution des descripteurs de forme dans ces sacs de mots par des valeurs provenant de l'information de couleur. Leurs approches sont similaires puisqu'ils proposent de travailler dans un espace à 5 dimensions correspondant aux 3 composantes couleur et aux 2 coordonnées spatiales de chaque pixel. Ils tentent alors de déterminer les paramètres du mélange de gaussiennes qui est le plus proche de la distribution des pixels dans cet espace 5D. Les probabilités d'appartenance à ces gaussiennes sont alors les poids utilisés dans les sacs de mots visuels fondés sur la forme.

La seconde méthode de combinaison séquentielle, proposée par Farag *et al.* [27] met en œuvre un classifieur Bayésien fondé sur l'information de teinte et est utilisé pour détecter dans une image les pixels qui appartiennent potentiellement à une classe d'objets donnés. Ainsi, plusieurs images d'objets peuvent être associées à une image après cette première étape utilisant exclusivement l'information de couleur. Ensuite, les descripteurs SIFT de l'image traitée sont extraits et com-

parés à ceux des objets de la base. La couleur est donc utilisée pour limiter le nombre d'objets à comparer avec l'image à traiter et la forme est utilisée pour affiner la décision finale. De la même manière, la méthode de Wu *et al* consiste à établir un premier appariement de régions entre deux images d'objets à comparer en utilisant la couleur [106] et à valider ou invalider ce choix par la comparaison des descripteurs SIFT de ces régions. Le descripteur couleur utilisé est composé des moyennes et écarts-types de la région considérée pour les composantes  $a^*b^*$  de l'espace  $L^*a^*b^*$ .

Les deux méthodes ci-dessous proposent plutôt d'utiliser la couleur dans un second temps après un appariement de points fondé sur la forme. Ainsi, Ancuti *et al.* considèrent que le nombre de points caractéristiques appariés après l'utilisation des SIFT est trop faible pour comparer deux images [11]. Ils proposent donc d'augmenter le nombre d'appariements en utilisant l'information de couleur. La première étape consiste donc à utiliser la méthode de Lowe pour appairer les points caractéristiques en utilisant l'information de forme [68]. Cependant, les critères de non-appariement étant trop stricts avec cette méthode d'après les auteurs, ces derniers proposent de reprendre les points non appariés et de tenter de les appairer avec l'information de couleur. Le descripteur couleur utilisé ici est l'histogramme des co-occurrences couleur proposé par Chang [18]. Cet histogramme compte le nombre co-occurrences de paires de couleur dans un même voisinage. Les auteurs montrent que l'utilisation d'un simple seuil sur l'intersection entre ces histogrammes locaux permet d'ajouter des appariements entre deux images sans augmenter le nombre de faux appariements. A la différence de Ancuti *et al.*, Goedeme *et al.* [50] estiment, au contraire, que la méthode d'appariement de Lowe conduit à un trop grand nombre de faux appariements et propose de valider chaque appariement fondé sur la forme par une prise en compte de la couleur. Le descripteur couleur choisi cette fois repose sur les moments couleur proposés par Mindru *et al.* [76].

### 2.4.3 Comparaisons parallèles

Nous allons maintenant évoquer les approches exploitant parallèlement descripteurs de forme et de couleur.

A l'instar de Van de Weijer *et al.* [101], Hegazy *et al.* [52] proposent d'associer le descripteur SIFT à l'histogramme des angles de composantes opposées (37-D) [99]. Cependant, plutôt que de concaténer ces deux descripteurs, les auteurs proposent de calculer les probabilités de présence d'un objet dans une image en considérant d'une part le descripteur SIFT et d'autre part le descripteur couleur. La décision finale est prise grâce au classifieur Adaboost qui combine ces deux probabilités. Des tests de classification sur les bases Caltech [1] et Graz02 [2] montrent que cette combinaison de descripteurs fournit toujours de meilleurs résultats que l'exploitation de chaque descripteur indépendamment.

Wu *et al.* [107] ont proposé, quant à eux, d'associer au descripteur SIFT, un histogramme couleur calculé dans l'espace HSV ( $6^3=216$ -D) pour suivre des objets. La méthode de suivi repose sur un filtre particulaire pour lequel les poids des particules sont mis à jour itérativement, en prenant en compte alternativement le descripteur couleur et le descripteur de forme. Les résultats de tracking montrent que la prise en compte simultanée des informations de couleur et de forme permet d'obtenir de meilleurs résultats que si seule la couleur est prise en compte.

Hu *et al.* [53] utilisent les auto-corrélogrammes comme descripteurs couleur. Un auto-corrélogramme est un tableau à double entrées dont un axe représente une couleur  $C_i$  et l'autre une distance  $d_j$  [54]. La cellule de coordonnées  $(i, j)$  représente le nombre de fois qu'un couple de pixels de couleur  $C_i$  se situe à une distance  $d_j$  l'un de l'autre. Ce descripteur représente donc l'évolution des paires de pixels de même couleur en fonction de la distance dans l'image. Ainsi, Hu *et al.* calculent la similarité entre deux objets en effectuant le rapport d'une mesure de similarité entre les ensembles de descripteurs SIFT des deux objets sur une distance entre leurs auto-corrélogrammes. La combinaison entre la forme et la

couleur a donc lieu au niveau du calcul de la distance entre deux objets.

Schügerl *et al.* [91] choisissent de combiner le descripteur SIFT avec un descripteur couleur recommandé dans le standard de compression vidéo MPEG-7. Ce descripteur couleur divise la région détectée en 64 blocs organisés sur une grille 8x8, chaque bloc étant caractérisé par la moyenne des pixels qui le constituent dans l'espace YCrCb. Une décomposition en cosinus discret (DCT) est alors appliquée sur chaque plan couleur et seuls les coefficients représentant les basses fréquences sont conservés pour décrire la région. Le descripteur est donc un ensemble de coefficients DCT calculés sur les plans Y, Cb et Cr. Pour une image requête donnée, les deux descripteurs de chaque région locale détectée vont être comparés aux ensembles de descripteurs des objets de la base et ainsi, chaque descripteur va voter, indépendamment de l'autre, pour un objet. L'image requête sera alors associée aux objets de la base ayant obtenu le plus grand nombre de votes.

Enfin, Nilsback *et al.* [87] combinent la couleur et la forme à partir de descripteur de forme utilisant les SIFT et les histogrammes de gradients tandis que le descripteur couleur est l'histogramme calculé dans l'espace HSV. Dans le cadre de la classification d'images de fleurs, Nilsback *et al.* montrent que la combinaison de la forme et de la couleur permet d'améliorer nettement les résultats fournis par chaque caractéristique indépendamment.

### 2.4.4 Descripteurs spatio-colorimétriques

Nous concluons maintenant cette partie par les approches spatio-colorimétriques.

Les méthodes de cette catégorie sont fondées sur l'extraction de descripteurs spatio-colorimétriques, c'est à dire de descripteurs donc chaque dimension contient à la fois une information de couleur et une information de forme. Les descripteurs spatio-colorimétriques les plus utilisés sont les SIFT couleur. Les SIFT ayant été initialement définis pour les images en niveaux de gris, leur structure ne convient pas à des images multi-composantes. Les SIFT couleur reposent donc simplement sur une généralisation des SIFT classiques appliqués marginalement à chaque

composante couleur. Par exemple, Bosch *et al.* [15] proposent d'extraire les SIFT dans l'espace couleur HSV et montrent que les résultats de reconnaissance sont meilleurs que ceux obtenus avec les SIFT extraits des images en niveaux de gris. L'originalité des travaux proposant des SIFT couleur reposent uniquement sur les composantes invariantes couleur utilisées pour extraire ces descripteurs. Le défaut de ces approches est que la dimension déjà élevée du descripteur SIFT (128) est multipliée par le nombre de composantes couleur. Pour extraire leurs descripteurs SIFT couleur Abdel-Hakim *et al.* [10] et Burghouts *et al.* [17] ont proposé d'utiliser les invariants couleur de Geusebroek *et al.* [45] qui reposent sur l'espace couleur gaussien (cf. page 50). Grâce à une étude comparative de différents descripteurs de niveaux de gris, de couleur, de forme et spatio-colorimétriques, Burghouts *et al.* montrent que le descripteur SIFT fondé sur la composante invariante C (cf. page 52) fournit les meilleurs résultats. Van de Sande *et al.* [98] ont aussi testé de nombreux descripteurs locaux couleur, de forme et spatio-colorimétriques et ils ont montré que les SIFT extraits de l'espace couleur opposé (cf. page 61) sont les plus performants en règle générale. Notons cependant que les descripteurs SIFT extraits de l'espace couleur gaussien n'ont pas été intégrés à ces tests. Finalement, Chu *et al.* [21] ont montré que les descripteurs en niveaux de gris SURF peuvent également être appliqués sur différents plans couleur et fournir de bons résultats. Ils montrent notamment que les descripteurs SURF fondés sur la composante invariante C (cf. page 52) fournit des résultats très légèrement supérieurs aux SURF extraits de l'espace couleur opposé  $o_1o_2o_3$ .

D'autres approches ont également été proposées pour extraire des descripteurs spatio-colorimétriques des images. Luke *et al.* s'inspirent de la structure des SIFT, qui est une concaténation d'histogrammes d'orientations de gradients pondérés par l'amplitude, proposent de caractériser le contenu d'une région locale par une concaténation d'histogrammes d'orientation de teintes pondérés par la saturation [69]. L'idée est ici que la teinte d'un pixel est d'autant plus significative que sa saturation est élevée. Les auteurs montrent que ce descripteur fournit de

meilleurs résultats que les SIFT en niveaux de gris. S’inspirant des travaux de Matas *et al.* [71], Naik *et al.* proposent de décrire une image par un ensemble de descripteurs locaux couleur [85]. Pour cela, ils détectent toutes les régions qui contiennent plusieurs couleurs et chaque région est caractérisée par l’ensemble des couleurs qu’elle contient. Les régions, ainsi extraites dans une même image, sont comparées deux à deux par une distance de Hausdorff de manière à ne conserver qu’un ensemble de régions distinctes. L’ensemble de descripteurs locaux qui caractérisent l’image représente donc la manière dont s’organisent les couleurs localement dans l’image. Geusebroek [46] a aussi proposé des descripteurs couleur locaux invariants qui sont fondés sur des dérivées spatiales calculées dans l’espace couleur gaussien [44]. En effet, il propose de caractériser chaque région par 12 paramètres de Weibull extraits des histogrammes de ces composantes dérivées et montre que ces descripteurs sont peu sensibles aux rotations de l’objet dans l’espace puisqu’à la différence des SIFT, ils ne sont pas fondés sur les directions de gradient dans l’espace image. La manière la plus intuitive de combiner localement la couleur et la forme consiste à réorganiser chaque région carrée de taille  $N \times N$  en un vecteur de dimension  $3N^2$  où le 3 correspond au nombre de composantes couleur qui caractérisent chaque pixel. Qiu [89] propose d’utiliser cette représentation avec l’espace couleur YCrCb. Ainsi, il sous-échantillonne toutes les régions carrées une imagerie de taille  $4 \times 4$  de laquelle il extrait l’information de luminance Y, donnant ainsi un vecteur 16-D achromatique, puis il sous-échantillonne à nouveau en une imagerie de taille  $2 \times 2$  dont il extrait les composantes de chrominance Cr et Cb, conduisant à deux vecteurs 4-D chromatiques. Il applique ensuite une quantification vectorielle dans les espaces de description 16-D et 4-D pour extraire les mots visuels chromatiques et achromatiques. Une image est alors caractérisée par les histogrammes de mots visuels (sacs de mots) chromatiques et achromatiques. Qiu montre que cette représentation fournit de meilleurs résultats que les auto-corrélogrammes couleur sur deux bases d’images couleur.



## 2.5 Conclusion

Le principale contribution de ce chapitre est qu'il dresse un état de l'art original des caractéristiques couleur invariantes notamment utilisées dans le cadre de la reconnaissance d'objets ou de l'indexation d'images. Nous avons montré, en nous référant à de nombreux travaux, que la couleur est une information essentielle pour reconnaître des objets car d'une part elle augmente le pouvoir discriminant des descripteurs, et d'autre part elle permet une invariance des descripteurs par rapport aux variations radiométriques et photométriques. Nous avons également montré que cette invariance doit être adaptée à l'application car toute contrainte additionnelle, toute hypothèse restrictive diminue le pouvoir discriminant. Dans ce chapitre, nous avons aussi montré de quelles manières la couleur peut être normalisée pour déterminer des caractéristiques insensibles à certaines variations photométriques ou radiométriques. Cette normalisation est fondée sur des hypothèses précises qui portent sur la formation de la couleur, les propriétés de réflexion des éléments de surface, les sensibilités des capteurs du système d'acquisition, les caractéristiques de l'éclairage ou directement sur le modèle de variations photométriques et radiométriques. Nous avons alors dressé une liste de caractéristiques couleur qui, sous ces hypothèses, permettent d'obtenir une insensibilité à telle ou telle variation.

Dans le cadre de cette étude, nous avons restreint notre champ d'analyse à une description générale la plus synthétique du domaine. Nous renvoyons le lecteur, intéressé par une description plus précise de telle ou telle caractéristique couleur invariantes, vers les références citées. Nous renvoyons également le lecteur, intéressé par une comparaison quantitative ou applicative des caractéristiques couleur invariantes vers ces mêmes références. Dans ce chapitre nous nous sommes focalisés uniquement sur des aspects fondamentaux, la plupart des caractéristiques couleur proposées étant basés sur des paramètres d'étude différents, des hypothèses pas nécessairement identiques, des propriétés d'invariance plus ou moins

## CHAPITRE 2. LES DESCRIPTEURS COULEUR INVARIANTS DE LA LITTÉRATURE

---

contraignantes (cf. telles 1.4 à 1.4), et des champs d'application dissimilaires.

Dans la seconde partie de ce chapitre, nous avons montré de quelle manière la couleur peut être plus ou moins prise en compte dans les descripteurs locaux classiques. Nous avons vu qu'une des tendances récentes est de l'associer au descripteur de forme SIFT. La combinaison peut être faite par une simple concaténation de descripteurs, par une prise en compte séquentielle de ces descripteurs, par un traitement parallèle ou directement par un descripteur spatio-colorimétrique dont chaque dimension est liée à ces deux informations.

Dans le cadre de cette seconde partie, nous avons également restreint notre champs d'analyse aux différentes possibilités de combiner un descripteur couleur à un descripteur de forme. Nous renvoyons le lecteur intéressé par les descripteurs locaux, autres que les SIFT (et par extension du SURF), vers les références du domaine. Dans ce chapitre nous nous sommes focalisés sur les SIFT car ce descripteur est le descripteur de référence dans le domaine pour combiner descripteurs couleurs et descripteurs de forme (cf Annexe B). Par la suite (cf. Chapitre 6), nous nous attacherons à effectuer des comparaisons via des tests comparatifs.

Les dernières études récentes, citées dans ce chapitre, ont montré que les SIFT couleur fournissaient les meilleurs résultats en terme d'appariement de régions et de classification d'objets. Nous introduirons donc ces descripteurs dans nos tests expérimentaux du chapitre 6.

Cependant, le problème de cette description orientée gradient local est double. D'une part, il s'agit d'une information très locale qui peut être sensible aux bruits et aux variations importantes de points de vue [44, 79]. De plus, comme nous le démontrerons par la suite, il est aussi très difficile de résumer cette somme d'informations très locales dans un descripteur compact. Les SIFT couleur ont en effet une dimension assez élevée (généralement 384), ce qui peut se révéler problématique dans le cadre de la classification pour laquelle des méthodes d'échantillonnage sont nécessaires (e.g. sacs de mots visuels). D'autre part, les descripteurs locaux qui utilisent cette information de gradient ont été initialement définis pour

les images à niveaux de gris dans lesquelles les pixels sont caractérisés par une seule valeur. Leur extension à la couleur consiste donc simplement à extraire indépendamment les descripteurs niveaux de gris de chaque composante couleur, que celles-ci soient invariantes ou non [10, 15, 17, 98]. La dimension vectorielle de la couleur n'est donc que partiellement prise en compte dans ces descripteurs. Pour toutes ces raisons, nous allons donc proposer une approche alternative dans le chapitre suivant.



---

---

## CHAPITRE 3

---

# INVARIANCE PAR TRANSFORMATIONS AFFINES ENTRE ESPACES COULEUR ET IMAGE

### Table des matières

---

<b>3.1</b>	<b>Introduction</b>	<b>83</b>
<b>3.2</b>	<b>Transformations affines entre deux espaces différents</b>	<b>84</b>
3.2.1	Solution des moindres carrés	85
3.2.2	Pouvoir discriminant	90
3.2.3	Invariance et covariance	92
3.2.4	Contraintes sur la transformation $T_{OVD}$	94
<b>3.3</b>	<b>De l'espace image vers l'espace couleur</b>	<b>99</b>
3.3.1	Illustration	99

## CHAPITRE 3. INVARIANCE PAR TRANSFORMATIONS AFFINES ENTRE ESPACES COULEUR ET IMAGE

---

3.3.2	Le descripteur IVC . . . . .	101
3.3.3	Les contraintes géométriques et couleur d'IVC . . . .	104
<b>3.4</b>	<b>De l'espace couleur à l'espace image . . . . .</b>	<b>108</b>
3.4.1	Illustration . . . . .	109
3.4.2	Le descripteur CVI . . . . .	109
<b>3.5</b>	<b>Conclusion . . . . .</b>	<b>113</b>

---

### 3.1 Introduction

Dans ce chapitre, nous introduisons deux nouveaux descripteurs couleur locaux. A l'inverse de la tendance actuelle qui consiste à utiliser l'information de gradients calculés en chaque pixel (GLOH [73], SURF [13], PCA-SIFT [57] ou SIFT couleur [10, 15, 17, 98]), nous proposons de résumer le contenu d'une région extraite d'une image en utilisant la répartition spatiale des couleurs dans cette région. En effet, l'utilisation de l'information de gradients locaux pose deux problèmes majeurs. Tout d'abord, comme cette information est très locale puisque calculée à l'échelle du pixel, (i) elle est sensible au bruit dans les images, (ii) comme les variations géométriques perturbent beaucoup l'organisation locale des pixels, cette information n'est que partiellement invariante aux transformations affines dans l'espace 3D de la scène observée (rotations 3D, détection approximative) [46, 79] et enfin (iii) il est difficile de résumer de manière compacte cette somme d'informations locales. Ce qui explique pourquoi la plupart des descripteurs qui utilisent le gradient ne sont pas compacts. Le second problème provient du fait que les descripteurs qui utilisent l'information de gradients ont été initialement définis pour caractériser le contenu d'images en niveaux de gris. Leur extension à la couleur consiste donc dans la plupart des cas à extraire le descripteur sur chaque composante de manière indépendante sans exploiter la structure vectorielle de la couleur [10, 15, 17, 98].

Ainsi, même si les SIFT couleur fournissent de très bons résultats en terme de reconnaissance et de classification d'objets [17, 98], nous défendons l'hypothèse qu'ils ne constituent pas la méthode la plus pertinente pour combiner l'information de couleur et de forme dans les images. Comme la plupart des systèmes d'acquisition fournissent aujourd'hui des images couleur, nous proposons, dans ce chapitre, des descripteurs locaux spécifiquement définis pour prendre en compte simultanément les 3 composantes couleur des pixels. En effet, chaque pixel étant caractérisé par 3 composantes dans l'espace couleur (*RVB*) et 2 dans

l'espace image ( $xy$ ), nous considérons l'ensemble des pixels présents dans une région détectée et estimons les 'meilleures' transformations affines qui permettent de passer des coordonnées spatiales aux coordonnées couleur et inversement. Bien entendu, ce type de transformation n'a aucune signification physique, mais nous montrons qu'elle permet d'obtenir des coordonnées qui dépendent des couleurs présentes dans la région considérée ainsi que de la manière dont ces couleurs sont organisées spatialement. Cette propriété leur procure un pouvoir discriminant important qui peut être exploité par un descripteur local. De plus, la transformation de l'espace image vers l'espace couleur (IVC) fournit des coordonnées qui sont invariantes à n'importe quelle transformation affine dans l'espace image alors que les coordonnées obtenues par transformation de l'espace couleur vers l'espace image (CVI) sont invariantes à toute transformation affine appliquée dans l'espace couleur. Ainsi, en plus d'être compacte, la concaténation du descripteur IVC et du descripteur CVI gère très bien le compromis entre pouvoir discriminant et invariance.

Le chapitre est organisé de la façon suivante. Dans la seconde partie, nous discutons les avantages inhérents à l'utilisation d'une transformation affine entre deux espaces, quels que soient ces espaces. Nous considérons ensuite la spécificité de chaque transformation dans les parties suivantes, à savoir la partie 3 sera consacrée à la transformation de l'espace image vers l'espace couleur alors que la partie 4 sera consacrée à la transformation de l'espace couleur vers l'espace image.

### **3.2 Transformations affines entre deux espaces différents**

Dans ce chapitre, nous allons expliquer comment appliquer des transformations affines d'un espace vers un autre. Comme nous travaillons à partir des informations couleur et spatiale associées aux pixels, l'un de ces espaces sera l'es-



pace 2D image dans lequel les pixels sont caractérisés par leurs coordonnées  $x$  et  $y$  et l'autre espace sera l'espace couleur dans lequel les coordonnées associées aux pixels seront les niveaux de rouge  $R$ , vert  $V$  et bleu  $B$ . Notre approche exploite deux transformations, l'une de l'espace image vers l'espace couleur (IVC) et l'autre de l'espace couleur vers l'espace image (CVI). Ces transformations ont des points communs qui vont être traités dans cette partie et des spécificités qui seront analysées dans les parties suivantes. L'objet de cette partie est d'expliquer la manière dont sont calculées les transformations entre deux espaces ainsi que les propriétés des coordonnées obtenues après application de ces transformations.

### 3.2.1 Solution des moindres carrés

Considérons une ellipse détectée autour d'un pixel dans une image et le rectangle de taille  $N \times M$  qui contient cette ellipse (cf. fig. 3.1). Chaque pixel  $P_i$ ,  $i \in [1, NM]$ , de ce rectangle est caractérisé par 2 coordonnées spatiales notées  $x_i$  et  $y_i$  ainsi que par 3 coordonnées couleur  $R_i$ ,  $V_i$  et  $B_i$ <sup>1</sup>. L'idée consiste à calculer la transformation affine d'un de ces espaces (image ou couleur) vers l'autre espace (couleur ou image, respectivement). Dans cette partie, de manière à généraliser la présentation, nous proposons de présenter la transformation d'un espace origine noté  $\mathcal{O}$  vers un espace destination noté  $\mathcal{D}$ , en gardant bien à l'esprit que ces espaces sont des espaces image ou couleur (cf. fig. 3.2).

La dimension de l'espace d'origine est notée  $N_{\mathcal{O}}$  et celle de l'espace destination  $N_{\mathcal{D}}$ , de sorte que  $N_{\mathcal{O}}$  et  $N_{\mathcal{D}} \in \{2, 3\}$ . Les coordonnées du pixel  $P_i$  sont  $(O_i^1, O_i^2, \dots, O_i^{N_{\mathcal{O}}})$  dans l'espace d'origine et  $(D_i^1, D_i^2, \dots, D_i^{N_{\mathcal{D}}})$  dans l'espace destination. L'objectif consiste à déterminer une transformation affine  $T_{\mathcal{O} \rightarrow \mathcal{D}}$  qui permet

---

<sup>1</sup>Pour alléger les notations, nous proposons de noter ces coordonnées  $R_i$ ,  $V_i$  et  $B_i$  plutôt que  $C^R(P_i)$ ,  $C^V(P_i)$  et  $C^B(P_i)$  comme dans le chapitre précédent

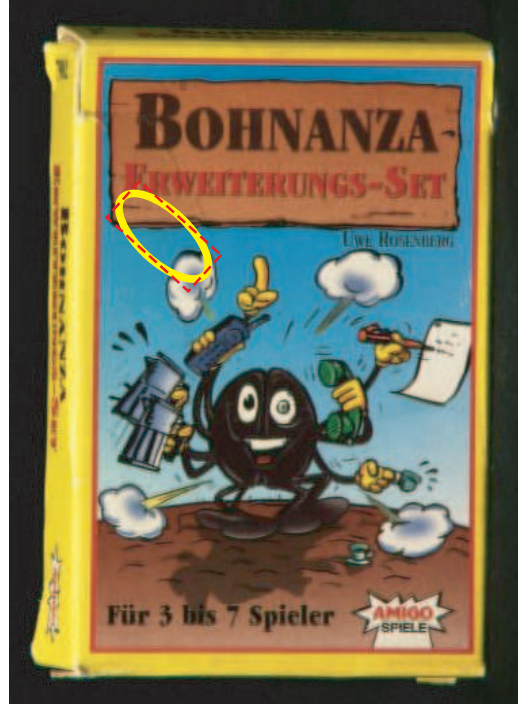


FIGURE 3.1 – Une ellipse détectée autour d’un pixel dans une image et le rectangle  $N \times M$  qui la contient.

le passage des coordonnées d’origine aux coordonnées de destination de telle sorte que :

$$\begin{bmatrix} D_i^1 & D_i^2 & \dots & D_i^{N_{\mathcal{D}}} \end{bmatrix} = \begin{bmatrix} O_i^1 & O_i^2 & \dots & O_i^{N_{\mathcal{O}}} & 1 \end{bmatrix} \cdot T_{OVD}, \quad (3.1)$$

où le coefficient 1 est introduit dans la matrice des coordonnées d’origine afin de prendre en compte une translation éventuelle de manière à ce que  $T_{OVD}$ , la matrice de transformation de taille  $(N_{\mathcal{O}} + 1) \times N_{\mathcal{D}}$  soit équivalente à une transformation affine.

Cette transformation peut facilement être déterminée si nous considérons un seul pixel mais lorsque les  $NM$  pixels de la région considérée sont pris en compte, nous pouvons simplement trouver une approximation de telle sorte que :

$$M_{\mathcal{D}} \approx M_{\mathcal{O}} \cdot T_{OVD}, \quad (3.2)$$

### CHAPITRE 3. INVARIANCE PAR TRANSFORMATIONS AFFINES ENTRE ESPACES COULEUR ET IMAGE

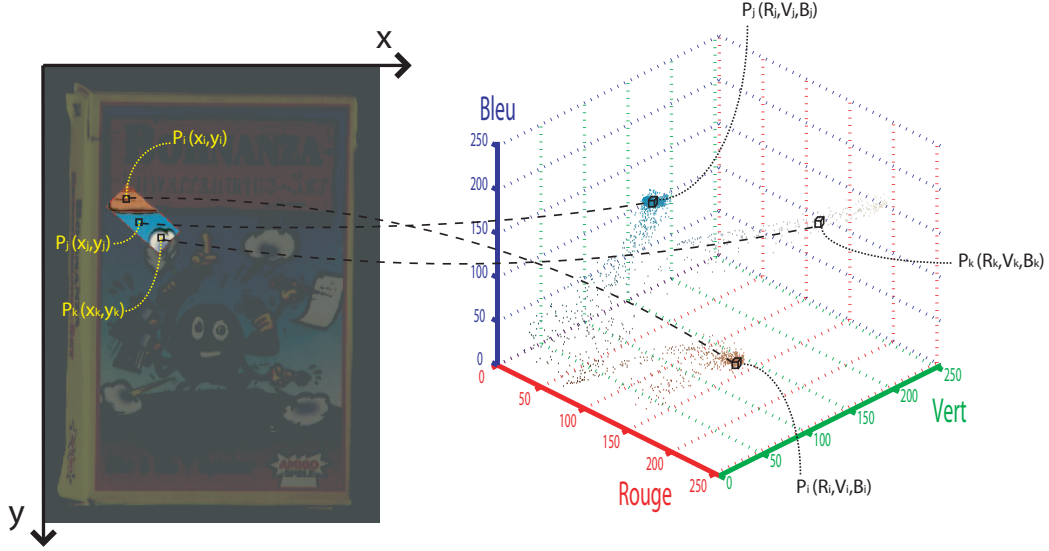


FIGURE 3.2 – L'espace image  $2D$  et l'espace couleur  $3D$ . L'idée consiste à déterminer la 'meilleure' transformation affine de l'espace image vers l'espace couleur (IVC) ou de l'espace couleur vers l'espace image (CVI). Pour IVC (CVI, resp.) l'espace origine  $\mathcal{O}$  est l'espace image (couleur, resp.) et l'espace destination  $\mathcal{D}$  est l'espace couleur (image, resp.).

où

$$M_{\mathcal{D}} = \begin{bmatrix} D_1^1 & D_1^2 & \dots & D_1^{N_{\mathcal{D}}} \\ D_2^1 & D_2^2 & \dots & D_2^{N_{\mathcal{D}}} \\ \vdots & \vdots & \ddots & \vdots \\ D_{NM}^1 & D_{NM}^2 & \dots & D_{NM}^{N_{\mathcal{D}}} \end{bmatrix} \text{ est une matrice } (NM) \times N_{\mathcal{D}}, \quad (3.3)$$

et

$$M_{\mathcal{O}} = \begin{bmatrix} O_1^1 & O_1^2 & \dots & O_1^{N_{\mathcal{O}}} & 1 \\ O_2^1 & O_2^2 & \dots & O_2^{N_{\mathcal{O}}} & 1 \\ & & \cdot & & \\ & & \cdot & & \\ & & \cdot & & \\ O_{NM}^1 & O_{NM}^2 & \dots & O_{NM}^{N_{\mathcal{O}}} & 1 \end{bmatrix} \text{ est une matrice } (NM) \times (N_{\mathcal{O}} + 1). \quad (3.4)$$

Pour estimer une solution de l'équation (3.2), nous proposons d'appliquer la méthode des moindres-carrés qui fournit la transformation  $T_{OVD}$  minimisant la somme des résidus au carré, c'est à dire la somme des distances entre les points à associer dans l'espace de destination. Cette solution est définie comme :

$$T_{OVD} = M_{\mathcal{O}}^+ M_{\mathcal{D}}, \quad (3.5)$$

où  $M_{\mathcal{O}}^+$  est la pseudo-inverse de la matrice  $M_{\mathcal{O}}$ .

Il existe plusieurs pseudo-inverses pour une matrice donnée, nous proposons d'utiliser celle fournie par la décomposition en valeurs singulières (SVD). La SVD décompose la matrice  $M_{\mathcal{O}}$  de taille  $NM \times (N_{\mathcal{O}} + 1)$  en une matrice  $U$  carrée de taille  $NM \times NM$ , une matrice  $S$  diagonale de taille  $NM \times (N_{\mathcal{O}} + 1)$  et une matrice carrée  $V$  de taille  $(N_{\mathcal{O}} + 1) \times (N_{\mathcal{O}} + 1)$  de telle sorte que :

$$M_{\mathcal{O}} = USV'. \quad (3.6)$$

De plus, les matrices  $U$ ,  $V$  et  $S$  ont les propriétés suivantes :

- $U$  est orthonormale, c'est à dire  $UU' = U'U = I_{NM}$ ,
- $V$  est orthonormale, c'est à dire  $VV' = V'V = I_{N_{\mathcal{O}}+1}$ ,
- $S^+S = Y_{k/(N_{\mathcal{O}}+1)}$  et  $SS^+ = Y_{k/NM}$ ,

où  $I_n$  est la matrice identité  $n \times n$ ,  $k = \text{rang}(M_{\mathcal{O}})$  et  $Y_{n/m}$ ,  $m > n$  est une matrice  $m \times m$  constituée de la matrice identité  $n \times n$  dans son coin haut-gauche et de zéros partout ailleurs :

$$Y_{n/m} = \begin{bmatrix} & 0 & \dots & 0 \\ & I_n & \vdots & \vdots \\ 0 & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & 0 \end{bmatrix}. \quad (3.7)$$

Une autre décomposition qui va être utile pour notre travail peut être déduite de la SVD. En effet, la matrice  $M_{\mathcal{O}}$  de rang  $k$  peut être obtenue par le produit de la matrice  $U_k$  de taille  $k \times k$ , de la matrice  $S_k$  de taille  $k \times k$  et de la transposée de la matrice  $V_k$  de taille  $(N_{\mathcal{O}} + 1) \times k$ , et ce de la manière suivante :

$$M_{\mathcal{O}} = U_k S_k V_k', \quad (3.8)$$

où les matrices  $U_k$ ,  $V_k$  and  $S_k$  ont les propriétés suivantes :

- $U_k' U_k = I_k$  mais  $U_k U_k' \neq I_k$ ,
- $V_k' V_k = I_k$  mais  $V_k V_k' \neq I_k$ ,
- $S_k^+ S_k = S_k S_k^+ = I_k$ .

Grâce à ces deux décompositions, la pseudo-inverse de la matrice  $M_{\mathcal{O}}$  est définie par :

$$M_{\mathcal{O}}^+ = V_k S_k^+ U_k' = V_k S_k^+ U_k'. \quad (3.9)$$

A partir de cette pseudo-inverse et de l'équation (3.5), nous pouvons déterminer la transformation  $T_{OVD}$  qui permet le passage de l'espace origine à l'espace destination. Dans les paragraphes suivants, nous présentons les propriétés d'une telle transformation.

### 3.2.2 Pouvoir discriminant

Suivant l'approche des moindres-carrés, nous avons obtenu une transformation affine  $T_{OVD}$ . L'application de cette transformation aux coordonnées d'origine  $(O_i^1, O_i^2, \dots, O_i^{N_\theta})$  du pixel  $P_i$  appartenant à la région considérée fournit des coordonnées dans l'espace de destination notées  $(\hat{D}_i^1, \hat{D}_i^2, \dots, \hat{D}_i^{\hat{N}_\theta})$  :

$$[O_i^1 O_i^2 \dots O_i^{N_\theta} 1] T_{OVD} = [\hat{D}_i^1 \hat{D}_i^2 \dots \hat{D}_i^{\hat{N}_\theta}]. \quad (3.10)$$

Puisqu'il n'y a aucune raison qu'une transformation affine existe entre l'espace image et l'espace couleur pour une image quelconque, les coordonnées  $(\hat{D}_i^1, \hat{D}_i^2, \dots, \hat{D}_i^{\hat{N}_\theta})$  sont différentes des coordonnées  $(D_i^1, D_i^2, \dots, D_i^{N_\theta})$ . En effet, nous pouvons noter que pour une région donnée, la seule manière d'avoir une transformation affine entre l'espace image et l'espace couleur est d'avoir une région qui représente des variations linéaires de couleurs proportionnelles aux coordonnées spatiales. La figure 3.3 représente le plan couleur qui s'ajuste au mieux à la distribution couleur du rectangle de la figure 3.2. Dans la figure 3.3, les points couleur représentent la distribution couleur de la région considérée, c'est à dire les coordonnées  $(D_i^1, D_i^2, \dots, D_i^{N_\theta})$  que nous cherchons à apparier avec la transformation des coordonnées image correspondantes et les points qui constituent le plan couleur sont les coordonnées  $(\hat{D}_i^1, \hat{D}_i^2, \dots, \hat{D}_i^{\hat{N}_\theta})$  effectivement obtenues après application de la transformation. Les lignes noires relient les paires de points correspondants de coordonnées  $\{\mathbf{D}_i, \hat{\mathbf{D}}_i\}$  de telle sorte que la longueur de ces lignes est directement liée aux valeurs de résidus. L'approche des moindres-carrés minimise la somme des résidus au carré, c'est à dire la somme des longueurs de ces lignes. Dans un souci de clarté, seule une partie des lignes est représentée.

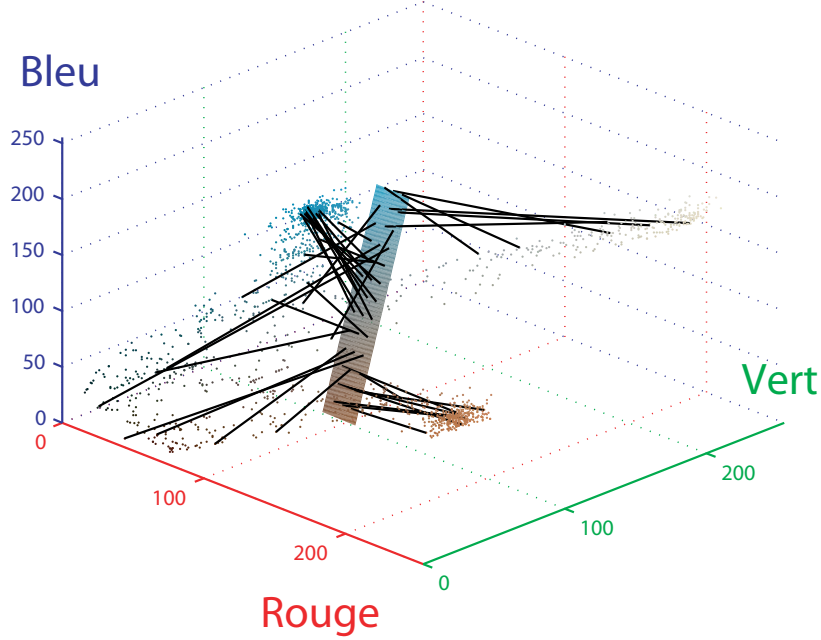


FIGURE 3.3 – Transformation de l’espace image vers l’espace couleur du rectangle détecté dans la figure 3.2. Se référer au texte pour les explications.

En combinant les équations (3.5) et (3.10), nous avons :

$$\begin{bmatrix} \hat{D}_i^1 & \hat{D}_i^2 & \dots & \hat{D}_i^{\hat{N}_{\mathcal{D}}} \end{bmatrix} = \begin{bmatrix} O_i^1 & O_i^2 & \dots & O_i^{N_{\mathcal{O}}} & 1 \end{bmatrix} M_{\mathcal{O}}^+ M_{\mathcal{D}}. \quad (3.11)$$

Nous pouvons donc noter que les coordonnées  $(\hat{D}_i^1, \hat{D}_i^2, \dots, \hat{D}_i^{\hat{N}_{\mathcal{D}}})$  d’un seul pixel  $P_i$  ne dépendent pas seulement de sa couleur et de sa position dans l’image mais aussi de toutes les couleurs présentes dans la région détectée et de leur distribution spatiale dans cette région. Ceci montre que, après l’application d’une telle transformation affine, quelques coordonnées dans l’espace de destination contiennent une quantité importante d’information. Le paragraphe suivant discute de la robustesse de ces coordonnées en cas de variations des conditions d’acquisition.

### 3.2.3 Invariance et covariance

Nous considérons un ensemble de points dans l'espace d'origine dont les coordonnées constituent la matrice  $M_{\mathcal{O}_1}$  et un autre ensemble de points dans l'espace de destination dont les coordonnées constituent la matrice  $M_{\mathcal{D}_1}$ . Chaque point dans l'espace d'origine est associé à un point dans l'espace de destination. L'objectif est d'évaluer la meilleure transformation affine globale  $T_{OVD_1}$  de l'espace d'origine vers l'espace destination. A partir des paragraphes précédents, nous savons que cette transformation  $T_{OVD_1}$  est approchée par :

$$T_{OVD_1} = M_{\mathcal{O}_1}^+ M_{\mathcal{D}_1}. \quad (3.12)$$

En appliquant cette transformation aux points dans l'espace d'origine, nous obtenons des points dans l'espace destination dont les coordonnées sont regroupées dans la matrice  $\hat{M}_{\mathcal{D}_1}$  de telle sorte que :

$$\begin{aligned} \hat{M}_{\mathcal{D}_1} &= M_{\mathcal{O}_1} T_{OVD_1}, \\ \hat{M}_{\mathcal{D}_1} &= M_{\mathcal{O}_1} M_{\mathcal{O}_1}^+ M_{\mathcal{D}_1}. \end{aligned} \quad (3.13)$$

En utilisant la SVD de la matrice  $M_{\mathcal{O}_1}$  obtenue par l'équation (3.8) ainsi que les propriétés des matrices utilisées pour cette décomposition, nous avons :

$$\begin{aligned} \hat{M}_{\mathcal{D}_1} &= M_{\mathcal{O}_1} M_{\mathcal{O}_1}^+ M_{\mathcal{D}_1}, \\ \hat{M}_{\mathcal{D}_1} &= U_{k_1} S_{k_1} V_{k_1}' V_{k_1} S_{k_1}^+ U_{k_1}' M_{\mathcal{D}_1}, \\ \hat{M}_{\mathcal{D}_1} &= U_{k_1} S_{k_1} I_{k_1} S_{k_1}^+ U_{k_1}' M_{\mathcal{D}_1}, \\ \hat{M}_{\mathcal{D}_1} &= U_{k_1} S_{k_1} S_{k_1}^+ U_{k_1}' M_{\mathcal{D}_1}, \\ \hat{M}_{\mathcal{D}_1} &= U_{k_1} I_{k_1} U_{k_1}' M_{\mathcal{D}_1}, \\ \hat{M}_{\mathcal{D}_1} &= U_{k_1} U_{k_1}' M_{\mathcal{D}_1}, \end{aligned} \quad (3.14)$$

où  $k_1 = \text{rang}(M_{\mathcal{O}_1})$ .

Considérons maintenant deux transformations affines  $T_O$  et  $T_D$  et appliquons



### CHAPITRE 3. INVARIANCE PAR TRANSFORMATIONS AFFINES ENTRE ESPACES COULEUR ET IMAGE

---

les respectivement aux ensembles de points dont les coordonnées constituent les matrices  $M_{\mathcal{O}_1}$  et  $M_{\mathcal{D}_1}$  :

$$M_{\mathcal{O}_2} = M_{\mathcal{O}_1} T_O, \quad (3.15)$$

et

$$M_{\mathcal{D}_2} = M_{\mathcal{D}_1} T_D. \quad (3.16)$$

Nous proposons de mesurer l'impact de ces transformations sur les coordonnées  $\hat{M}_{\mathcal{D}_2}$  dans l'espace de destination. Nous posons au préalable une restriction sur les transformations  $T_O$  et  $T_D$  puisque nous supposons que  $\text{rang}(M_{\mathcal{O}_1} T_O) = \text{rang}(M_{\mathcal{O}_1})$  et que  $\text{rang}(M_{\mathcal{D}_1} T_D) = \text{rang}(M_{\mathcal{D}_1})$ , c'est à dire que ces transformations ne diminuent pas le nombre de dimensions indépendantes dans les espaces considérés. Par exemple, si nous sommes dans l'espace image et que les points considérés ne sont pas alignés, l'application de la transformation ne doit pas retourner des points alignés. De même, le nombre de composantes couleur indépendantes ne doit pas diminuer par application de la transformation.

A partir de ces deux matrices  $M_{\mathcal{O}_2}$  et  $M_{\mathcal{D}_2}$ , nous pouvons déduire la matrice  $\hat{M}_{\mathcal{D}_2}$  :

$$\begin{aligned} \hat{M}_{\mathcal{D}_2} &= M_{\mathcal{O}_2} M_{\mathcal{O}_2}^+ M_{\mathcal{D}_2}, \\ \hat{M}_{\mathcal{D}_2} &= M_{\mathcal{O}_1} T_O (M_{\mathcal{O}_1} T_O)^+ M_{\mathcal{D}_1} T_D, \\ \hat{M}_{\mathcal{D}_2} &= U_{k_1} S_{k_1} V_{k_1}' T_O (U_{k_1} S_{k_1} V_{k_1}' T_O)^+ M_{\mathcal{D}_1} T_D. \end{aligned}$$

Comme  $U_{k_1}' U_{k_1} = I_{k_1}$  (cf. eq. (3.8)), nous obtenons ensuite :

$$\begin{aligned} \hat{M}_{\mathcal{D}_2} &= U_{k_1} S_{k_1} V_{k_1}' T_O (S_{k_1} V_{k_1}' T_O)^+ U_{k_1}^+ M_{\mathcal{D}_1} T_D, \\ \hat{M}_{\mathcal{D}_2} &= U_{k_1} S_{k_1} V_{k_1}' T_O (S_{k_1} V_{k_1}' T_O)^+ U_{k_1}' M_{\mathcal{D}_1} T_D, \end{aligned}$$

et comme  $S_{k_1} V_{k_1}' T_O$  est une matrice  $k_1 \times N_{\mathcal{O}_1}$  dont le rang est  $k_1$  et  $k_1 \leq N_{\mathcal{O}_1}$ , nous avons :

$$S_{k_1} V_{k_1}' T_O (S_{k_1} V_{k_1}' T_O)^+ = I_{k_1} \quad (3.17)$$

et donc :

$$\hat{M}_{\mathcal{D}_2} = U_{k_1} U_{k_1}' M_{\mathcal{D}_1} T_D. \quad (3.18)$$

En utilisant l'équation (3.14), nous déduisons que  $\hat{M}_{\mathcal{D}_2} = \hat{M}_{\mathcal{D}_1} T_D$ . Comme ce résultat ne dépend pas de  $T_O$ , nous en déduisons que les coordonnées  $\hat{M}_{\mathcal{D}_2}$  sont invariantes à toute transformation affine  $T_O$  appliquée au préalable dans l'espace d'origine. De plus, nous notons aussi que ces coordonnées sont covariantes (c'est à dire qu'elles varient de la même manière) à toute transformation affine appliquée au préalable dans l'espace de destination. Ceci est illustré dans la figure 3.4 qui présente trois cas de transformation (un par ligne). Dans un souci de clarté, les espace origine et destination sont des espaces  $2D$ . La première colonne représente les points dans l'espace origine, la seconde les points dans l'espace de destination et la troisième le résultat, c'est à dire les positions des points de l'espace origine après leur avoir appliqué la transformation affine  $T_{OVD}$ . Nous remarquons que quand une transformation affine est appliquée au préalable dans l'espace d'origine (seconde ligne), les coordonnées résultantes dans l'espace de destination ne sont pas modifiées alors que lorsqu'une transformation affine est appliquée dans l'espace de destination au préalable (troisième ligne), les coordonnées résultantes sont aussi modifiées de la même manière. Ceci montre les propriétés d'invariance et de covariance de notre approche.

### 3.2.4 Contraintes sur la transformation $T_{OVD}$

Nous avons souligné dans le chapitre 2 que toute invariance superflue pouvait conduire à une baisse du pouvoir discriminant et il pourrait donc paraître opportun de contraindre la transformation affine  $T_{OVD}$  en fonction des variations qui sont susceptibles d'avoir lieu dans l'espace d'origine, de sorte à ne pas ajouter une invariance à une transformation qui ne peut pas se produire. Par exemple, pour une transformation de l'espace image vers l'espace couleur, si nous savons

que les objets dans les images ne sont sujets qu'à des changements d'échelle, des rotations dans le plan image et des translations, il n'est pas nécessaire d'introduire des paramètres de cisaillement dans la transformation affine recherchée. Cependant, nous montrons ci-dessous que l'application d'une contrainte sur la transformation  $T_{OVD}$  fait perdre aux coordonnées résultantes leur propriété de covariance avec une transformation appliquée au préalable dans l'espace destination (de type  $T_D$ ). Nous testerons dans le chapitre 6 une approche avec contrainte et une autre sans contrainte de manière à mesurer en pratique les avantages et les inconvénients de chacune. L'intégration de contraintes géométriques est détaillée dans la section 6.4.1

Pour notre démonstration, considérons le cas ci-dessus pour lequel nous avons deux configurations différentes, l'une caractérisée par les matrices  $M_{\mathcal{O}_1}$  et  $M_{\mathcal{D}_1}$  et l'autre par les matrices  $M_{\mathcal{O}_2}$  et  $M_{\mathcal{D}_2}$ , de telle sorte qu'il existe une transformation  $T_O$  entre les deux ensembles de coordonnées d'origine et une transformation  $T_D$  entre les ensembles de coordonnées de destination. Nous avons montré que sans utiliser de contraintes sur  $T_{OVD}$ , nous avons  $\hat{M}_{\mathcal{D}_2} = \hat{M}_{\mathcal{D}_1} T_D$ . Nous proposons de décomposer chacune des transformations  $T_{OVD_1}$  et  $T_{OVD_2}$  en un produit de matrices  $a_{OVD_i}$  et  $b_{OVD_i}$  telles que  $T_{OVD_1} = a_{OVD_1} b_{OVD_1}$  et  $T_{OVD_2} = a_{OVD_2} b_{OVD_2}$ . La matrice  $a_{OVD_i}$  représente donc une partie de la transformation (rotation et translation par exemple) alors que la matrice  $b_{OVD_i}$  représente le reste de la transformation (changement d'échelle et cisaillement, par exemple). Imposer des contraintes à la transformation  $T_{OVD_i}$  revient à considérer que la matrice de transformation est constituée d'une seule des deux matrices  $a_{OVD_i}$  et  $b_{OVD_i}$ . Nous notons  $\hat{T}_{OVD_i}$  la matrice de transformation contrainte tel que  $\hat{T}_{OVD_i} = a_{OVD_i}$ . A partir de là, nous avons :

$$T_{OVD_i} = \hat{T}_{OVD_i} b_{OVD_i}. \quad (3.19)$$

Nous en déduisons alors :

$$\begin{aligned}
 \hat{M}_{\mathcal{D}_1} &= M_{\mathcal{O}_1} \hat{T}_{OVD_1}, \\
 \hat{M}_{\mathcal{D}_1} b_{OVD_1} &= M_{\mathcal{O}_1} \hat{T}_{OVD_1} b_{OVD_1}, \\
 \hat{M}_{\mathcal{D}_1} b_{OVD_1} &= M_{\mathcal{O}_1} T_{OVD_1}, \\
 \hat{M}_{\mathcal{D}_1} b_{OVD_1} &= M_{\mathcal{O}_1} M_{\mathcal{O}_1}^+ M_{\mathcal{D}_1}, \\
 \hat{M}_{\mathcal{D}_1} b_{OVD_1} &= U_{k_1} U_{k_1}' M_{\mathcal{D}_1}.
 \end{aligned} \tag{3.20}$$

De la même manière nous obtenons :

$$\begin{aligned}
 \hat{M}_{\mathcal{D}_2} &= M_{\mathcal{O}_2} \hat{T}_{OVD_2}, \\
 \hat{M}_{\mathcal{D}_2} b_{OVD_2} &= M_{\mathcal{O}_2} \hat{T}_{OVD_2} b_{OVD_2}, \\
 \hat{M}_{\mathcal{D}_2} b_{OVD_2} &= M_{\mathcal{O}_2} M_{\mathcal{O}_2}^+ M_{\mathcal{D}_2}, \\
 \hat{M}_{\mathcal{D}_2} b_{OVD_2} &= M_{\mathcal{O}_1} T_O (M_{\mathcal{O}_1} T_O)^+ M_{\mathcal{D}_1} T_D, \\
 \hat{M}_{\mathcal{D}_2} b_{OVD_2} &= U_{k_1} U_{k_1}' M_{\mathcal{D}_1} T_D, \\
 \hat{M}_{\mathcal{D}_2} b_{OVD_2} &= \hat{M}_{\mathcal{D}_1} b_{OVD_1} T_D.
 \end{aligned} \tag{3.21}$$

Ce résultat montre que la propriété de covariance n'est plus vérifiée lorsque la transformation affine  $T_{OVD}$  est contrainte puisque nous n'obtenons plus  $\hat{M}_{\mathcal{D}_2} = \hat{D}_{\mathcal{D}_1} T_{\mathcal{D}}$ . Or, par la suite, nous allons montrer que cette propriété de covariance est exploitée lorsqu'il s'agit de contrôler le degré d'invariance du descripteur final par rapport aux variations dans l'espace de destination.

Dans cette partie, nous avons montré que l'utilisation de la méthode des moindres-carrés pour estimer la transformation  $T_{OVD}$  permettait d'extraire de manière compacte une information sur les couleurs présentes dans la région considérée ainsi que leur distribution spatiale. De plus, nous avons montré qu'après application de cette transformation, nous obtenons des coordonnées qui sont invariantes à toute transformation affine appliquée au préalable dans l'espace d'origine et qui sont covariantes aux transformations appliquées au préalable dans l'espace de destination, à condition que la transformation  $T_{OVD}$  ne soit pas contrainte. Les deux parties suivantes discutent les spécificités de ces deux transformations : de l'es-

### CHAPITRE 3. INVARIANCE PAR TRANSFORMATIONS AFFINES ENTRE ESPACES COULEUR ET IMAGE

---

pace image vers l'espace couleur et de l'espace couleur vers l'espace image.

### CHAPITRE 3. INVARIANCE PAR TRANSFORMATIONS AFFINES ENTRE ESPACES COULEUR ET IMAGE

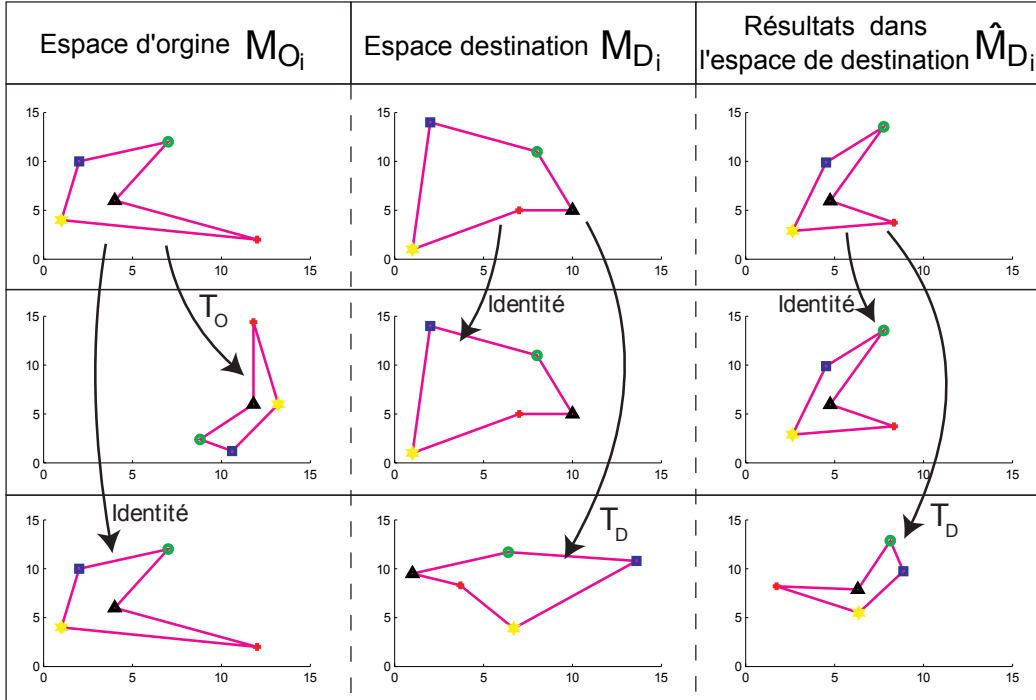


FIGURE 3.4 – Trois exemples de propriété d'invariance d'une transformation affine  $T_{OVD}$ . A chaque point correspond un seul et unique symbole dans les espaces origine et destination. A partir de ces paires, nous évaluons la transformation affine qui permet de passer des positions d'origine (première colonne) aux positions destination (seconde colonne). La troisième colonne montre les positions des points pris dans l'espace origine après leur avoir appliqué notre transformation  $T_{OVD}$ . Chaque ligne expose un cas. La seconde ligne reprend le cas de la première ligne après avoir appliqué une transformation  $T_O$  aux positions d'origine alors que pour la troisième ligne une transformation  $T_D$  est appliquée préalablement aux positions destination. Nous remarquons dans la troisième colonne que les positions résultantes sont invariantes à  $T_O$  et covariantes avec  $T_D$ .

### 3.3 De l'espace image vers l'espace couleur

Dans cette partie, nous considérons que l'espace origine est l'espace image  $\mathcal{I}$  et que l'espace destination est l'espace couleur  $RVB \mathcal{C}$ .

#### 3.3.1 Illustration

En suivant l'approche présentée dans la section précédente, nous calculons la transformation affine  $T_{IVC}$  de l'espace image vers l'espace couleur puis nous appliquons cette transformation aux coordonnées spatiales dans l'image notées  $xy$ . Les coordonnées 3D résultantes notées  $R\hat{V}B$  sont exprimées dans l'espace couleur  $RVB$ . Nous avons déjà montré, d'une part, que les coordonnées  $R\hat{V}B$  d'un seul pixel sont liées aux couleurs de tous les pixels présents dans la région considérée ainsi qu'à leur distribution spatiale et d'autre part, que ces coordonnées sont invariantes à toute transformation  $T_{xy}$  appliquée au préalable dans l'espace image et covariantes à toute transformation  $T_{RVB}$  appliquée au préalable dans l'espace couleur. Ceci est illustré dans la figure 3.5 qui présente trois configurations, une par ligne. La première colonne représente 3 images couleur dans lesquelles une région est détectée. Ces régions décrivent grossièrement les mêmes surfaces d'étude par rapport à un objet donné. Dans la seconde colonne, les pixels qui vont être pris en compte pour calculer la transformation  $T_{IVC}$  sont mis en valeur. La troisième colonne donne une idée de la nouvelle répartition des pixels dans l'espace couleur après leur avoir appliqué la transformation  $T_{IVC}$ . L'image de la seconde ligne est obtenue par application d'une transformation  $T_{xy}$  aux coordonnées spatiales des pixels de l'image de la première ligne alors que celle de la troisième ligne est obtenue par application d'une transformation  $T_{RVB}$  aux couleurs des pixels de l'image de la première ligne. Les positions des pixels dans la troisième colonne illustrent leurs propriétés d'invariance à  $T_{xy}$  et de covariance à  $T_{RVB}$ . En effet, les positions des première et deuxième lignes (troisième colonne) sont très similaires alors qu'il existe une transformation affine entre les coordonnées spatiales

### CHAPITRE 3. INVARIANCE PAR TRANSFORMATIONS AFFINES ENTRE ESPACES COULEUR ET IMAGE

des images initiales correspondantes (première colonne). La légère différence qui existe entre les coordonnées  $R\hat{V}B$  sont dues aux petites erreurs de détection en terme de position, échelle et orientation. Enfin, nous notons que la transformation affine  $T_{RVB}$  appliquée aux couleurs entre les images des première et troisième lignes est aussi celle qui relie les coordonnées  $R\hat{V}B$  des pixels des mêmes lignes dans la troisième colonne.

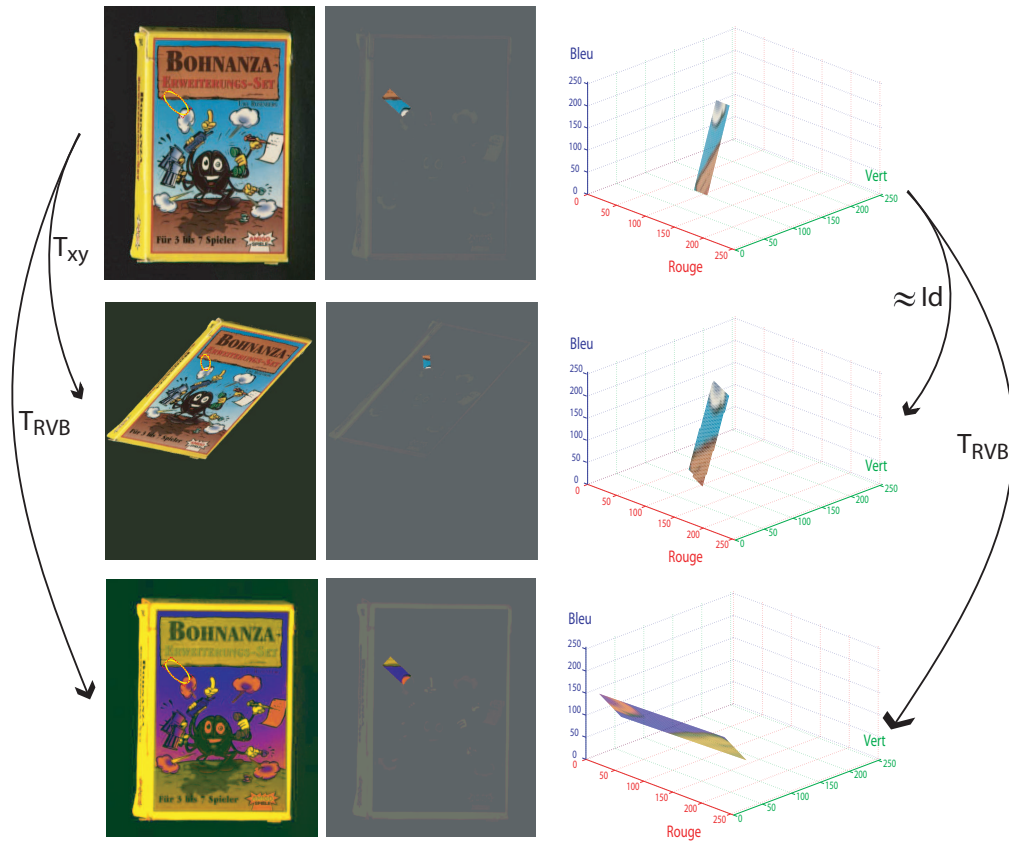


FIGURE 3.5 – Invariance et covariance des coordonnées  $R\hat{V}B$  obtenues par application de la transformation  $T_{IVC}$ . Voir le texte pour les explications.



### 3.3.2 Le descripteur IVC

Le descripteur local couleur IVC est extrait des coordonnées couleur  $R\hat{V}B$  obtenues par application de la transformation  $T_{IVC}$ . Comme nous considérons des régions rectangulaires dans l'espace image, les formes obtenues après transformation sont des parallélogrammes (cf. troisième colonne de la figure 3.5) dont les positions peuvent être exactement définies par la position de trois coins, c'est à dire par 9 coordonnées. Ces 9 valeurs pourraient constituer le descripteur final mais, de sorte à augmenter le pouvoir discriminant, plutôt que d'appliquer une transformation  $T_{IVC}$  à tous les pixels de la région détectée notée  $r_0$ , nous proposons de décomposer le rectangle en 4 sous rectangles de taille identique ( $r_1, r_2, r_3$  et  $r_4$ ) et d'appliquer indépendamment une transformation  $T_{IVC}$  à chacun d'entre eux. Nous obtenons ainsi 4 parallélogrammes notés  $\hat{r}_1, \hat{r}_2, \hat{r}_3$  et  $\hat{r}_4$  dont les positions peuvent être définies par  $4 \times 9 = 36$  valeurs. 36 sera donc la dimension de ce descripteur.

Ce descripteur est invariant à toute transformation affine appliquée dans l'espace image, mais il est très sensible aux variations photométriques qui perturbent les couleurs des pixels. Nous proposons donc de normaliser les coordonnées des parallélogrammes  $\hat{r}_i$  afin d'obtenir une invariance contrôlée, c'est à dire dont le degré peut être adapté à la base d'images utilisée. Ce degré d'invariance sera un paramètre du descripteur. Pour cela, nous exploitons la propriété de covariance des coordonnées des coins des parallélogrammes  $\hat{r}_i$  avec les transformations appliquées dans l'espace couleur. En effet, puisque ces coordonnées sont définies dans l'espace couleur, elles peuvent être normalisées comme les composantes couleur. Ainsi, de par la propriété de covariance, si une variation photométrique dans l'espace couleur peut être modélisée par une transformation affine (comme nous l'avons montré dans le chapitre 2), nous savons que nous pouvons supprimer l'effet de cette variation en appliquant l'inverse de cette transformation aux coordonnées des parallélogrammes  $\hat{r}_i$ . Comme nous travaillons sous des condi-

tions d'acquisition non contrôlées, cette transformation n'est pas connue. Nous proposons donc plutôt de normaliser ces coordonnées en appliquant une transformation affine  $T_{norm}$  dont le nombre de degrés de liberté (ddl) est contrôlé. Ainsi, en fonction du degré d'invariance souhaité, la transformation  $T_{norm}$  sera une matrice diagonale (3 ddl dans le modèle diagonal de l'éq. (2.20)), ou une matrice diagonale et une matrice de translation  $3 \times 1$  (6 ddl dans le modèle diagonal et translation de l'éq. (2.21)), ou une matrice  $3 \times 3$  (9 ddl dans le modèle linéaire de l'éq. (2.22)), ou une transformation affine 3D (12 ddl dans le modèle affine de l'éq. (2.23)). Les méthodes de calcul de la transformation  $T_{norm}$  en fonction du modèle choisi sont présentées dans la partie 3.3.3. Comme l'ont fait Mindru *et al.* [78], nous allons considérer ici les modèles à 3, 6 et 12 ddl uniquement. Lorsque le modèle des fonctions monotones croissantes (paragraphe 2.2.5) est retenu, la normalisation (égalisations d'histogrammes 1D) est alors plutôt appliquée comme un prétraitement à la région concernée.

Pour déterminer cette transformation de normalisation  $T_{norm}$ , nous proposons d'utiliser le parallélogramme  $\hat{r}_0$  obtenu en appliquant la transformation  $T_{IVC}$  au rectangle  $r_0$ . Cependant, les coordonnées de ce parallélogramme ne sont pas intégrées dans le descripteur final, elles sont simplement utilisées pour normaliser celles des parallélogrammes  $\hat{r}_1$ ,  $\hat{r}_2$ ,  $\hat{r}_3$  et  $\hat{r}_4$ . Pour cela, nous proposons de fixer arbitrairement 4 points canoniques qui forment un parallélogramme dans l'espace couleur. Le choix des positions absolues de ces points n'a que peu d'importance car elles sont fixées une fois pour toute et sont les mêmes pour toutes les régions comparées. Ensuite, utilisant une nouvelle fois l'approche des moindres carrés, nous évaluons la meilleure transformation  $T_{norm}$  qui déplace les coins du parallélogramme  $\hat{r}_0$  au plus proche de ces positions canoniques. Cette transformation est alors appliquée aux parallélogrammes  $\hat{r}_1$ ,  $\hat{r}_2$ ,  $\hat{r}_3$  et  $\hat{r}_4$ . Après application de cette normalisation, les positions de 3 coins de ces 4 parallélogrammes constituent le descripteur 36D IVC normalisé.

Ce processus est illustré dans la figure 3.6 sur l'ellipse correspondant à celle

définie dans la figure 3.2. Tout d'abord, une transformation  $T_{IVC}$  (une différente pour chaque) est indépendamment appliquée à chaque rectangle  $r_1$ ,  $r_2$ ,  $r_3$  et  $r_4$  ainsi qu'au rectangle  $r_0$ , fournissant ainsi les 5 parallélogrammes  $\hat{r}_1$ ,  $\hat{r}_2$ ,  $\hat{r}_3$ ,  $\hat{r}_4$  et  $\hat{r}_0$  respectivement dans l'espace couleur. Ensuite la transformation  $T_{norm}$  qui déplacent les coins du parallélogramme  $\hat{r}_0$  au plus proche des positions canoniques (4 cercles bleus) est déterminée (approche des moindres-carrés) puis appliquée aux parallélogrammes  $\hat{r}_1$ ,  $\hat{r}_2$ ,  $\hat{r}_3$  et  $\hat{r}_4$ . Le descripteur IVC normalisé est constitué des coordonnées de 3 coins (points rouges) de chacun des 4 parallélogrammes dans l'espace couleur 3D représentés en bas à droite de la figure.

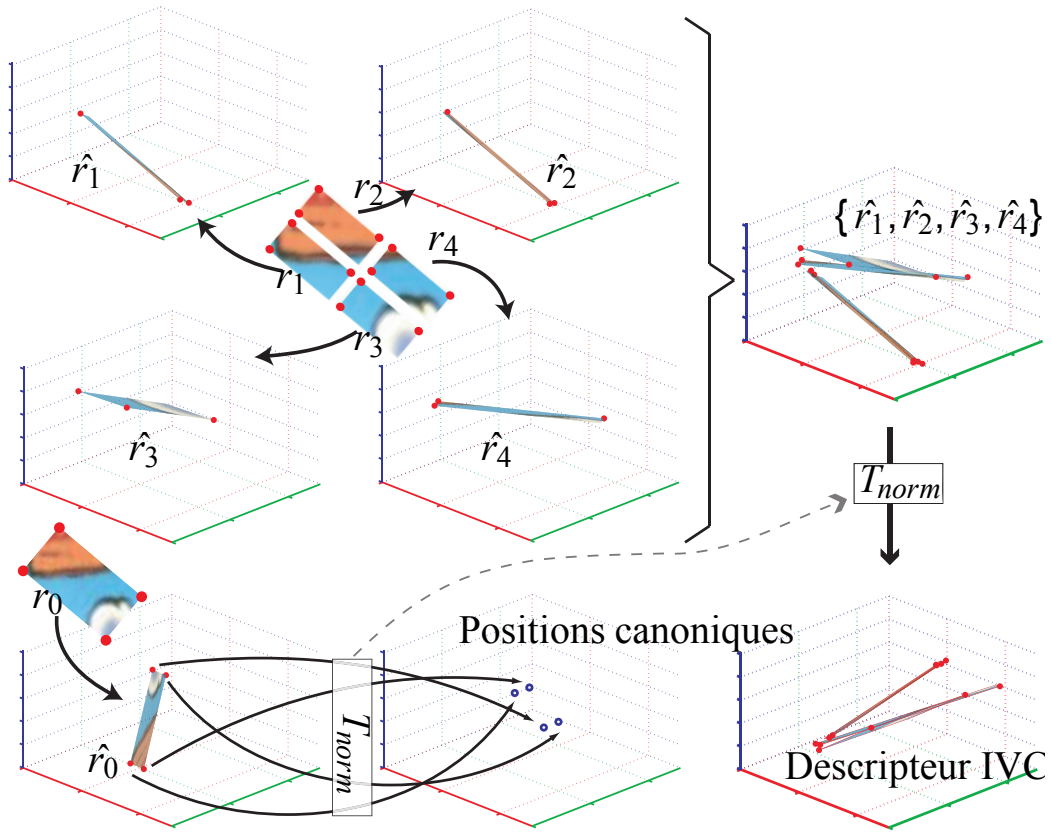


FIGURE 3.6 – Le descripteur IVC est constitué des coordonnées 3D des 12 coins rouges . Cf. le texte pour plus de détails.

### 3.3.3 Les contraintes géométriques et couleur d'IVC

Ce paragraphe expose de quelles manières les contraintes géométriques et couleur peuvent être imposées aux transformations utilisées pour déterminer le descripteur IVC.

#### Les contraintes géométriques

Dans le paragraphe 3.2.4, nous avons introduit la possibilité de contraindre la transformation  $T_{IVC}$  de manière à réduire le degré d'invariance géométrique du descripteur final pour augmenter son pouvoir discriminant. En effet, il est habituel de supposer que la transformation géométrique entre deux images peut être modélisée par une rotation dans le plan image, un changement d'échelle et une translation, négligeant ainsi le cisaillement. Nous allons montrer dans ce paragraphe comment nous pouvons imposer cette contrainte géométrique à la transformation recherchée. Cependant, nous avons souligné que la mise en place d'une telle contrainte annule la propriété de covariance des coordonnées obtenues par rapport aux variations photométriques. Il est donc indispensable d'utiliser cette contrainte lorsque la normalisation photométrique décrite dans le paragraphe précédent n'est pas nécessaire. De manière à pouvoir imposer une telle contrainte, il est nécessaire de considérer un espace de destination couleur à 2 dimensions, négligeant ainsi la 3<sup>ème</sup>. Dans ce cas, il est préférable de calculer 3 descripteurs indépendamment, chacun dans un plan couleur  $2D : \{RV, RB, VB\}$ . Nous allons ici présenter le cas du plan  $RV$  qui peut être facilement étendu aux 2 autres plans. Ainsi, chaque pixel  $P_i$  est caractérisé par 2 composantes spatiales  $x_i$  et  $y_i$  et 2 composantes couleur  $R_i$  et  $V_i$ . Dans ce cas précis, la transformation  $T_{IVC}$  peut être décomposée en une matrice  $2 \times 2$  et une matrice de translation  $2 \times 1$ . L'équation 3.1 peut alors être ré-écrite de la manière suivante :

$$\begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = \begin{bmatrix} R_i \\ V_i \end{bmatrix}, \quad (3.22)$$

### CHAPITRE 3. INVARIANCE PAR TRANSFORMATIONS AFFINES ENTRE ESPACES COULEUR ET IMAGE

---

où  $t_1$  et  $t_2$  sont les paramètres de translation et les coefficients  $m_j$ ,  $j = 1, \dots, 4$  sont les paramètres de rotation, changement d'échelle et cisaillement.

Lorsque le cisaillement peut être négligé, nous avons  $m_3 = -m_2$  et  $m_4 = m_1$  de telle sorte que :

$$\begin{bmatrix} m_1 & m_2 \\ -m_2 & m_1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = \begin{bmatrix} R_i \\ V_i \end{bmatrix}. \quad (3.23)$$

En regroupant les 4 paramètres recherchés dans une même matrice, nous obtenons :

$$\begin{bmatrix} x_i & y_i & 1 & 0 \\ y_i & -x_i & 0 & 1 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ t_1 \\ t_2 \end{bmatrix} = \begin{bmatrix} R_i \\ V_i \end{bmatrix}. \quad (3.24)$$

Cette équation ne prend en compte qu'un seul pixel et l'ajout des coordonnées des autres pixels de la région détectée s'effectue en ajoutant des lignes dans les matrices de part et d'autre :

$$\begin{bmatrix} x_i & y_i & 1 & 0 \\ y_i & -x_i & 0 & 1 \\ \dots & & & \\ \dots & & & \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ t_1 \\ t_2 \end{bmatrix} \approx \begin{bmatrix} R_i \\ V_i \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}. \quad (3.25)$$

Cette équation peut être ré-écrite comme :  $[xy]T \approx [RV]$ . La transformation  $T$  est alors déterminée par  $T = [xy]^+[RV]$ . Les 4 paramètres extraits de cette matrice  $4 \times 1$  sont les  $m_1$ ,  $m_2$ ,  $t_1$  et  $t_2$  recherchés qui vont constituer la matrice contrainte  $T_{IVC}$ .

### Les contraintes couleur

Nous allons ici expliquer de quelle manière les contraintes peuvent être imposées dans le calcul de la transformation  $T_{norm}$  qui permet d'obtenir une invariance contrôlée en normalisant les positions des coins du parallélogramme  $\hat{r}_0$ . En effet, cette transformation est celle qui déplace ces coins au plus proche des points canoniques (cf. Page 102). Tous ces points se situant dans l'espace couleur, nous allons noter leurs coordonnées  $R_i, V_i, B_i$  pour le  $i^{ème}$  coin du parallélogramme et  $R'_i, V'_i, B'_i$  pour le  $i^{ème}$  point canonique. L'idée étant de déplacer les 4 coins du parallélogramme vers les 4 points canoniques correspondants (le  $i^{ème}$  coin vers le  $i^{ème}$  point) avec la transformation  $T_{norm}$ , ce qui donne :

$$\begin{bmatrix} R'_1 & V'_1 & B'_1 \\ R'_2 & V'_2 & B'_2 \\ R'_3 & V'_3 & B'_3 \\ R'_4 & V'_4 & B'_4 \end{bmatrix} \approx \begin{bmatrix} R_1 & V_1 & B_1 & 1 \\ R_2 & V_2 & B_2 & 1 \\ R_3 & V_3 & B_3 & 1 \\ R_4 & V_4 & B_4 & 1 \end{bmatrix} [T_{norm}]. \quad (3.26)$$

Lorsque nous limitons la normalisation à 3 ddl (modèle diagonal de l'éq.(2.20)), nous obtenons alors :

$$\begin{bmatrix} R'_1 & V'_1 & B'_1 \\ R'_2 & V'_2 & B'_2 \\ R'_3 & V'_3 & B'_3 \\ R'_4 & V'_4 & B'_4 \end{bmatrix} \approx \begin{bmatrix} R_1 & V_1 & B_1 \\ R_2 & V_2 & B_2 \\ R_3 & V_3 & B_3 \\ R_4 & V_4 & B_4 \end{bmatrix} \begin{bmatrix} a_R & 0 & 0 \\ 0 & a_V & 0 \\ 0 & 0 & a_B \end{bmatrix}, \quad (3.27)$$

qui conduit à :

$$\begin{bmatrix} R'_1 \\ V'_1 \\ B'_1 \\ R'_2 \\ V'_2 \\ B'_2 \\ R'_3 \\ V'_3 \\ B'_3 \\ R'_4 \\ V'_4 \\ B'_4 \end{bmatrix} \approx \begin{bmatrix} R_1 & 0 & 0 \\ 0 & V_1 & 0 \\ 0 & 0 & B_1 \\ R_2 & 0 & 0 \\ 0 & V_2 & 0 \\ 0 & 0 & B_2 \\ R_3 & 0 & 0 \\ 0 & V_3 & 0 \\ 0 & 0 & B_3 \\ R_4 & 0 & 0 \\ 0 & V_4 & 0 \\ 0 & 0 & B_4 \end{bmatrix} \begin{bmatrix} a_R \\ a_V \\ a_B \end{bmatrix}. \quad (3.28)$$

La solution au sens des moindres carrés de cette équation nous permet alors de définir les 3 coefficients  $a_R$ ,  $a_V$  et  $a_B$ .

De la même manière, si nous optons pour la normalisation à 6 ddl (modèle diagonal et translation de l'éq.(2.21)), nous obtenons alors :

$$\begin{bmatrix} R'_1 & V'_1 & B'_1 \\ R'_2 & V'_2 & B'_2 \\ R'_3 & V'_3 & B'_3 \\ R'_4 & V'_4 & B'_4 \end{bmatrix} \approx \begin{bmatrix} R_1 & V_1 & B_1 \\ R_2 & V_2 & B_2 \\ R_3 & V_3 & B_3 \\ R_4 & V_4 & B_4 \end{bmatrix} \begin{bmatrix} a_R & 0 & 0 \\ 0 & a_V & 0 \\ 0 & 0 & a_B \end{bmatrix} + \begin{bmatrix} b_R \\ b_V \\ b_B \end{bmatrix}, \quad (3.29)$$

ce qui peut être ré-écrit de la manière suivante pour isoler les inconnues dans une

seule matrice :

$$\begin{bmatrix} R'_1 \\ V'_1 \\ B'_1 \\ R'_2 \\ V'_2 \\ B'_2 \\ R'_3 \\ V'_3 \\ B'_3 \\ R'_4 \\ V'_4 \\ B'_4 \end{bmatrix} \approx \begin{bmatrix} R_1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & V_1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & B_1 & 1 \\ R_2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & V_2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & B_2 & 1 \\ R_3 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & V_3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & B_3 & 1 \\ R_4 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & V_4 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & B_4 & 1 \end{bmatrix} \begin{bmatrix} a_R \\ b_R \\ a_V \\ b_V \\ a_B \\ b_B \end{bmatrix}. \quad (3.30)$$

Cette équation nous permet aussi de déterminer les 6 coefficients de la transformation.

Pour le modèle affine (équ.(2.23)) à 12 ddl, il n'y a aucune contrainte et les 12 paramètres de la transformation  $T_{norm}$  sont directement obtenus à partir de l'équation 3.26. Cependant, lorsque les normalisations des 3 composantes R, V et B sont indépendantes, comme pour le modèle affine, les 4 points à normaliser ne doivent plus être coplanaires, il faut donc prendre 3 coins du parallélogramme  $\hat{r}_0$  et 1 coin de  $\hat{r}_1$ .

Nous allons maintenant présenter le descripteur obtenu par transformation de l'espace couleur vers l'espace image.

### 3.4 De l'espace couleur à l'espace image

Dans cette partie, nous considérons que l'espace d'origine est l'espace couleur  $RVB \mathcal{C}$  et que l'espace de destination est l'espace image  $\mathcal{I}$ .



### 3.4.1 Illustration

L'objectif ici est d'évaluer la transformation  $T_{CVI}$  au sens des moindres-carrés depuis les coordonnées  $RVB$  des pixels jusqu'à leurs coordonnées spatiales  $xy$  et d'appliquer cette transformation aux coordonnées  $RVB$ . Les coordonnées résultantes  $\hat{x}\hat{y}$  d'un seul pixel sont aussi liées à la fois aux couleurs présentes dans la région considérée et à leur distribution spatiale. De plus, comme ces coordonnées sont invariantes à toute transformation affine appliquée au préalable dans l'espace couleur, elles sont insensibles à la plupart des variations photométriques.

La figure 3.7 présente les mêmes cas que ceux affichés dans la figure 3.5. Dans un souci de clarté, nous avons réduit la fenêtre d'observation des images aux ellipses détectées. Ainsi, la première colonne représente les ellipses détectées dans les images de la première colonne de la figure 3.5. Dans la seconde colonne sont affichées les distributions couleur des régions rectangulaires correspondantes alors que la troisième colonne montre les points couleur de ces distributions après leur avoir appliqué la transformation  $T_{CVI}$ . Ces représentations ne sont là que pour illustrer les distributions relatives des points couleur dans l'espace image. En effet, plusieurs points peuvent avoir les mêmes coordonnées dans cet espace et nous représentons ici uniquement le dernier point transformé à chaque position. Nous pouvons remarquer que les coordonnées  $\hat{x}\hat{y}$  obtenues après transformation sont covariantes avec la transformation affine spatiale  $T_{xy}$  et invariantes dans le cas d'une transformation affine couleur  $T_{RVB}$ . Ce sont bien les coordonnées des pixels dans l'espace image qui sont invariantes puisqu'elles ne changent pas après application d'une transformation couleur mais bien entendu les couleurs des pixels ont varié.

### 3.4.2 Le descripteur CVI

Le descripteur CVI est fondé sur les coordonnées  $\hat{x}\hat{y}$  obtenues par application de la transformation  $T_{CVI}$  à la distribution couleur de la région considérée.

### CHAPITRE 3. INVARIANCE PAR TRANSFORMATIONS AFFINES ENTRE ESPACES COULEUR ET IMAGE

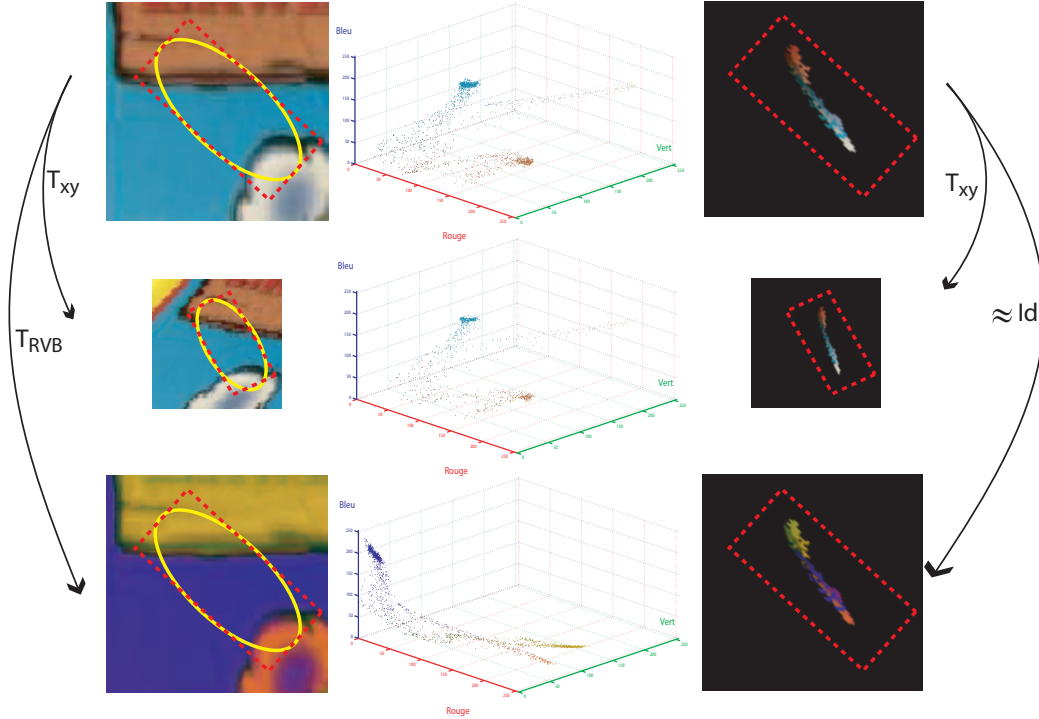


FIGURE 3.7 – Invariance et covariance des coordonnées  $\hat{x}\hat{y}$  obtenues après application de la transformation  $T_{CVI}$ . Cf. le texte pour plus de détails.

Ces coordonnées ne sont pas sensibles aux transformations affines couleur mais sont covariantes avec les transformations affines spatiales. Par conséquent, plutôt que de considérer les coordonnées absolues  $\hat{x}\hat{y}$  dans l'image, nous prenons en compte les coordonnées  $x_r\hat{y}_r$  relatives au rectangle associé à l'ellipse détectée. C'est à dire que l'origine du repère est le centre du rectangle et que les axes de coordonnées sont ses côtés dont les longueurs sont normalisées à 1, de sorte que  $\{x_r, y_r\} \in [-0.5, 0.5]$ . De par la propriété de covariance des coordonnées absolues, ces coordonnées relatives sont invariantes aux variations d'échelle, aux rotations et aux translations dans l'espace image.

Alors que les formes obtenues après application des transformations  $T_{IVC}$  aux rectangles détectés sont des parallélogrammes et peuvent donc être facilement résumées par quelques valeurs, les formes obtenues après application des trans-

formations  $T_{CVI}$  sont complexes car liées aux distributions couleur des régions concernées. Elles peuvent donc être difficilement résumées par quelques valeurs. Nous proposons donc d'exploiter quelques propriétés de l'approche des moindres-carrés pour caractériser l'ensemble des coordonnées  $x_r\hat{y}_r$ . En effet, puisque ces coordonnées sont le résultat de l'application d'une transformation affine dont le but est de minimiser les écarts avec les coordonnées  $x_r y_r$ , la transformation  $T_{x_r y_r V x_r \hat{y}_r}$  qui transforme les coordonnées relatives d'origine  $x_r y_r$  (première colonne de la figure 3.7) vers ces coordonnées  $x_r \hat{y}_r$  (troisième colonne de la figure 3.7) peut être modélisée par une matrice  $2 \times 2$  symétrique. Cette matrice décrit de quelle manière les coordonnées des pixels ont varié après l'application de la transformation  $T_{CVI}$ . Elle est donc porteuse d'information très intéressante pour décrire la distribution des coordonnées  $x_r \hat{y}_r$ . Par conséquent, nous proposons d'utiliser les trois valeurs différentes extraites de cette matrice  $2 \times 2$  symétrique pour constituer ce descripteur. Pour être concret, nous créons deux matrices  $NM \times 2$  : la matrice  $I_{x_r y_r}$  qui contient les coordonnées relatives des pixels d'origine (i.e.  $x_r y_r$ ) et la matrice  $I_{x_r \hat{y}_r}$  qui contient leur coordonnées relatives après application de la transformation  $T_{CVI}$ ,  $N \times M$  étant la taille de la région rectangulaire considérée. Ensuite, la transformation  $T_{x_r y_r V x_r \hat{y}_r}$  est définie comme :  $T_{x_r y_r V x_r \hat{y}_r} = I_{x_r y_r}^+ I_{x_r \hat{y}_r}$ .

Ainsi, les 3 valeurs extraites de la transformation  $T_{x_r y_r V x_r \hat{y}_r}$  sont invariantes à la plupart des variations géométriques et photométriques. De manière à augmenter le pouvoir discriminant du descripteur final CVI, nous proposons de décomposer le rectangle original en 16 sous-rectangles de taille identique, comme le font les descripteurs SIFT, et d'évaluer indépendamment ces trois valeurs pour chaque rectangle. Nous obtenons donc un descripteur de taille  $3 \times 16 = 48$ . Le processus est illustré dans la figure 3.8. Dans cette figure, l'image en haut à gauche représente le rectangle qui a été détecté dans l'image 3.1 puis, décomposé en 16 sous-rectangles. Une transformation  $T_{CVI}$  est calculée indépendamment pour chacun de ces sous-rectangles, comme exposée précédemment, et appliquée au sous-rectangle concerné de sorte à obtenir une nouvelle position pour chaque pixel

comme illustré dans l'image en haut à droite de la figure. Ainsi, à chaque pixel nous pouvons associer d'une part des coordonnées  $2D$   $x_r y_r$  (cf. l'image en haut à gauche dans la Figure 3.8) et d'autre part des coordonnées  $2D$   $x_r \hat{y}_r$  (cf. l'image en haut à droite dans la Figure 3.8). L'étape suivante consiste à calculer la transformation au sens des moindres-carrés, pour chaque rectangle indépendamment, pour passer des coordonnées  $x_r y_r$  aux coordonnées  $x_r \hat{y}_r$ . Ces 16 transformations sont illustrées par les flèches rouges dans la figure.

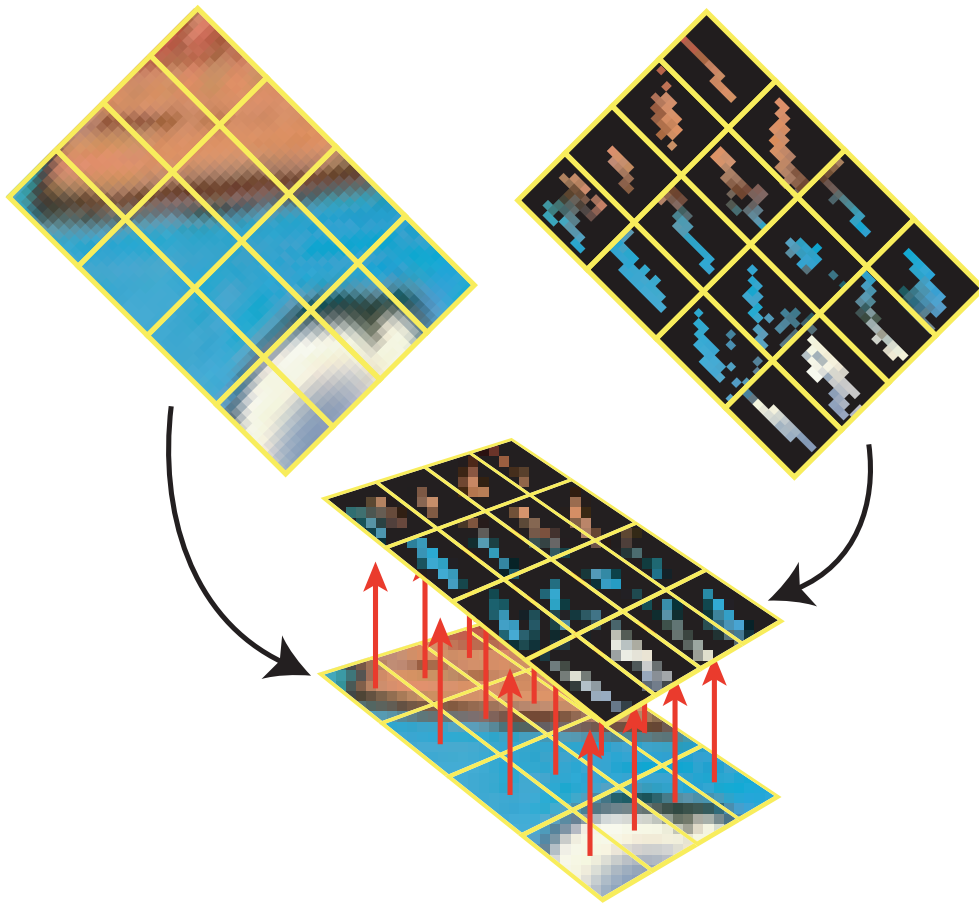


FIGURE 3.8 – Le descripteur CVI est constitué de  $3 \times 16$  valeurs extraites des 16 transformations  $2D$  représentées par les flèches rouges. Cf. le texte pour plus de détails.

### 3.5 Conclusion

Dans ce chapitre, nous avons introduit deux nouveaux descripteurs couleur qui représentent les distributions spatiales des couleurs dans les régions considérées. Pour cela, nous avons proposé de calculer puis d'appliquer des transformations affines entre l'espace image et l'espace couleur. Cette approche permet d'obtenir des descripteurs compacts, discriminants et présentant certaines propriétés d'invariance géométrique et photométrique. La transformation de l'espace image vers l'espace couleur nous permet de définir le descripteur à 36 dimensions noté *IVC* et la transformation de l'espace couleur à l'espace image conduit au descripteur à 48 dimensions noté *CVI*. Nous avons montré que le descripteur *IVC* est invariant à toute transformation affine appliquée dans l'espace image et que son degré d'invariance aux variations photométriques peut être adapté en fonction de l'application. Nous avons aussi montré que le descripteur *CVI* est invariant à toute transformation affine appliquée dans l'espace couleur, aux variations d'échelle, aux rotations dans le plan et aux translations dans l'espace image.

Chacun de ces descripteurs présente donc des informations importantes sur la région considérée avec des propriétés d'invariance géométrique et photométrique différentes. Lorsque les variations qui peuvent avoir lieu entre deux images d'un même objet sont connues, l'un ou l'autre de ces descripteurs peut être choisi en fonction des ces propriétés. Cependant, lorsque ces variations ne sont pas prévisibles, il est préférable de combiner ces deux descripteurs compacts avec une simple concaténation et de pondérer chaque dimension par l'écart-type qui la caractérise. Les performances de ces descripteurs pris individuellement et de leur concaténation seront décrites et analysées dans le chapitre 6.

Dans ce chapitre, nous avons considéré qu'une variation des conditions d'acquisition pouvait être modélisée par une transformation affine, nous nous sommes appuyés pour justifier cette hypothèse sur diverses études qui ont été présentées dans le chapitre 2.

### CHAPITRE 3. INVARIANCE PAR TRANSFORMATIONS AFFINES ENTRE ESPACES COULEUR ET IMAGE

---

Lorsque la transformation associée aux variations des conditions d'éclairage n'est pas linéaire, il est alors préférable d'utiliser les mesures de rangs des pixels. Dans le cadre de nos travaux, nous avons proposé un autre descripteur local qui exploite cette propriété, lequel sera est présenté dans le chapitre suivant.

---

---

## CHAPITRE 4

---

# L'INVARIANCE PAR MESURES DE RANGS

### Table des matières

---

4.1	Introduction . . . . .	116
4.2	Mesures de rangs . . . . .	117
4.3	Coefficients de corrélation de rangs de Kendall . . . . .	118
4.4	Corrélation de rangs entre sous-ensembles . . . . .	121
4.5	Le descripteur local fondé sur les corrélations de rang cou- leur . . . . .	126
4.6	Conclusion . . . . .	127

---

## 4.1 Introduction

Dans le chapitre précédent, nous avons considéré que la variation des conditions d'acquisition pouvait être modélisée par des transformations affines. Cette idée a été largement soutenue dans la littérature [25, 30, 33, 41, 48, 51, 75, 77]. Cependant, Finlayson *et al.* ont montré que cette modélisation atteint ses limites dans certains cas, notamment lorsque les capteurs d'acquisition varient, et ont donc proposé une transformation non-linéaire pour normaliser les composantes couleur [28]. L'idée consiste à ordonner les pixels en considérant chaque composante couleur indépendamment et à caractériser chaque pixel par les 3 positions obtenues dans les images de composantes rouge, verte et bleue suite à ces classements. Ces positions sont appelées 'mesures de rang'. Leur calcul est détaillé dans la seconde partie de ce chapitre. Une fois que l'on a associé à chaque pixel ses 3 mesures de rang, Finlayson *et al.* proposent de caractériser l'image ou la région considérée par l'histogramme 3D de ces mesures de rang. Cette approche a 3 inconvénients :

- chaque composante est traitée indépendamment dans le calcul des mesures de rang, ce qui néglige la dimension vectorielle de la couleur,
- le descripteur final ne prend pas en compte les positions des pixels dans les images ou dans les régions,
- la mesure de rang ne prend pas en compte le phénomène de métamérisme qui signifie que 2 surfaces élémentaires caractérisées par des couleurs différentes sous une condition d'illumination donnée peuvent être caractérisées par des couleurs identiques sous d'autres conditions d'illumination.

Dans ce chapitre, nous proposons donc un nouveau descripteur local qui à la fois n'est pas sensible aux variations non-linéaires des couleurs et ne souffre pas des inconvénients précédents. En effet, plutôt que d'utiliser directement les me-



mesures de rang des pixels dans les régions considérées, nous calculons des corrélations de mesures de rangs entre sous-ensembles de pixels. Comme les positions de ces sous-ensemble de pixels sont prises en compte, le descripteur final contient de fait une information spatiale. De plus, nous verrons dans la partie suivante que la mesure de corrélation utilisée (celle proposée par Kendall [58]) présente l'avantage de pouvoir prendre en compte le phénomène de métamérisme. Enfin, en plus des corrélations de rangs pour une même composante, nous allons aussi calculer les corrélations de rangs entre composantes différentes. La seconde partie de ce chapitre introduit les mesures de rangs.

## 4.2 Mesures de rangs

Pour définir les mesures de rangs, Finlayson *et al.* proposent de décomposer une image couleur en 3 images de composante rouge, verte et bleue dans lesquelles chaque pixel  $P_i$  est caractérisé par une seule valeur correspondant à son niveau de rouge  $R_i$ , vert  $V_i$  et bleu  $B_i$  respectivement. Ensuite, dans chacune de ces images de composante, les pixels sont triés dans l'ordre croissant de leur niveau et sont alors caractérisés par leur mesure de rangs :

$$\mathcal{M}^k[\mathbf{I}](P_i) = \frac{\text{Card}\{P_j \in \mathbf{I} / c^k(P_j) \leq c^k(P_i)\}}{\text{Card}\{P_j \in \mathbf{I}\}}. \quad (4.1)$$

Finlayson *et al.* ont prouvé que ces mesures de rangs sont peu sensibles aux changements de conditions d'acquisition telles que celles liées à l'illuminant ou aux capteurs et proposent donc de caractériser les pixels par leurs mesures de rang dans les trois images de composante plutôt que par leurs coordonnées selon chaque composante couleur. Cette transformation des composantes couleur en mesures de rangs est obtenue par une simple égalisation d'histogramme 1D de chacune des composantes. Finlayson *et al.* ont montré que cette égalisation fournit de très bons résultats de reconnaissance d'objets sur la base de l'Université de

l'East Anglia (UEA) par rapport à de nombreuses autres normalisations qui ont été présentées dans notre chapitre 2 telles que le greyworld [16], la normalisation MaxRGB [64], la normalisation itérative 'comprehensive normalization' [37], les rapports de niveaux de composante [40], les angles couleur [31], les moments invariants [49], les composantes  $c_1c_2c_3$ ,  $l_1l_2l_3$  ou  $m_1m_2m_3$  [47]. Muselet *et al.* ont également montré que ces mesures de rangs ne sont que partiellement préservées en cas de variations de changement d'éclairage. Ils ont donc proposé d'utiliser des mesures de rangs floues [82] ainsi que des matrices de co-occurrences adaptées fondées sur les mesures de rangs [83].

Étant donnés les bons résultats fournis par ces différentes approches, nous avons introduit les mesures de rangs dans notre descripteur local en exploitant des coefficients de corrélation de mesures de rangs tels que ceux fournis par Kendall [58].

### 4.3 Coefficients de corrélation de rangs de Kendall

Les coefficients de corrélation de rangs fournissent une information sur la dépendance entre deux variables en analysant la corrélation entre les mesures de rangs de ces variables. Lorsque l'on considère des images en couleur, les variables sont les niveaux des composantes couleur des pixels et les coefficients représentent donc les corrélations entre les mesures de rangs des intensités des pixels.

A cet effet, nous considérons une image couleur qui contient  $n_{pix}$  pixels  $P_i$ . Pour la mesure de corrélation, toutes les paires de composantes vont être prises en compte. Pour illustrer nos propos nous avons considéré les composantes rouge et verte, cela dit la même approche peut être appliquée à toutes les paires de composantes.

La corrélation de rangs de Kendall [58] nécessite d'analyser chaque paire de pixels  $\{P_i, P_j\}$ ,  $i \neq j$  d'une image couleur. Si les niveaux de l'intensité rouge et

verte des deux pixels  $P_i$  et  $P_j$  sont classés dans le même ordre, c'est à dire si  $R_i < R_j$  et  $V_i < V_j$  ou si  $R_i > R_j$  et  $V_i > V_j$ , la paire de  $\{P_i, P_j\}$  est appelée concordante. Sinon, si ces pixels sont de telle sorte que  $R_i < R_j$  et  $V_i > V_j$  ou que  $R_i > R_j$  et  $V_i < V_j$ , la paire est appelée discordante. Une fois que toutes les paires de pixels ont été analysées, nous évaluons la différence entre le nombre de paires concordantes et le nombre de paires discordantes. Cette différence est notée  $S$ . Pour évaluer le coefficient de Kendall  $\tau_{RV}$ , il faut simplement diviser cette différence  $S$  par le nombre de paires de pixels considérées  $\frac{n_{pix}(n_{pix}-1)}{2}$  de sorte que :

$$\tau_{RV} = \frac{2S}{n_{pix}(n_{pix} - 1)}. \quad (4.2)$$

Malheureusement, ce coefficient de corrélation de rangs ne prend pas en compte les paires de pixels dont les niveaux sont identiques ( $R_i = R_j$  ou  $V_i = V_j$ ). Dans ce cas, la paire n'est ni concordante ni discordante, elle n'est donc pas prise en compte lors de l'évaluation de  $S$ . Par conséquent, le  $\tau_{RV}$  de Kendall varie lorsque le nombre de paires de pixels caractérisés par le même niveau varie. Ceci pose d'autant plus de problèmes que Muselet *et al.* ont montré que les mesures de rangs des pixels ne sont pas exactement préservées en cas de changement d'éclairage [84]. En effet, le phénomène de métamérisme explique que certaines surfaces élémentaires peuvent avoir la même couleur sous un éclairage et des couleurs différentes sous un autre éclairage. La figure 4.1 empruntée à [80] montre ce phénomène. Par voie de conséquence, deux pixels qui ont les mêmes mesures de rangs sous un éclairage peuvent avoir des mesures de rangs différents sous un autre éclairage. Dans un tel cas d'étude, le coefficient de Kendall n'est pas constant.

Par conséquent, le coefficient de Kendall  $\tau_{RV}$  n'est que partiellement invariant aux variations d'éclairage. Heureusement, il existe une version corrigée  $\tau'_{RV}$  du coefficient de Kendall qui permet d'intégrer cette instabilité du nombre de pixels

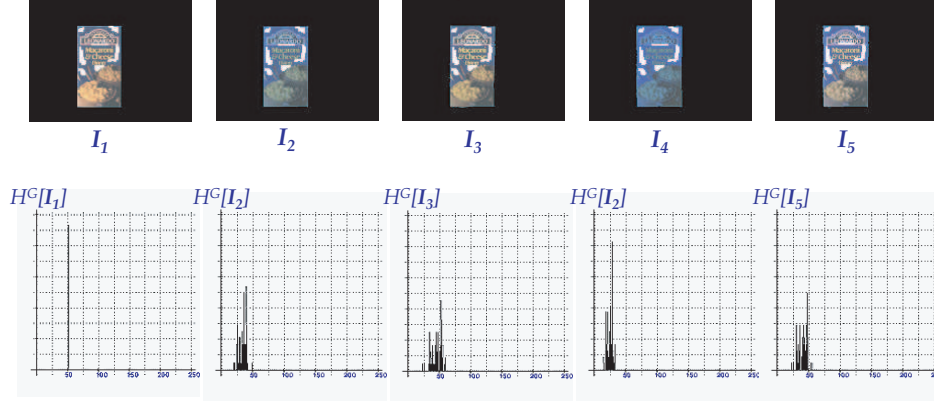


FIGURE 4.1 – Les pixels caractérisés par un niveau d'intensité de vert égal à 50 dans la première image sont présentés en blanc. La seconde ligne représente pour chaque image, l'histogramme vert de ces pixels représentés en blanc. Nous observons dans ces histogrammes que tous ces pixels n'ont pas tous le même niveau mais que leurs niveaux sont répartis autour d'une valeur. Ceci est expliqué pour une part par le bruit d'acquisition mais aussi et surtout par le phénomène de métamérisme.

caractérisés par des niveaux identiques :

$$\tau'_{RV} = \frac{S}{\sqrt{(\frac{1}{2}n_{pix}(n_{pix} - 1) - Tk_r)}} \times \frac{1}{\sqrt{(\frac{1}{2}n_{pix}(n_{pix} - 1) - Tk_v)}}, \quad (4.3)$$

où  $Tk_r = \frac{1}{2} \sum_{r=0}^{L-1} t_r(t_r - 1)$  et  $Tk_v = \frac{1}{2} \sum_{v=0}^{L-1} t_v(t_v - 1)$ .  $t_r$  ( $t_v$ , resp.) est le nombre de pixels caractérisés par un même niveau de rouge  $r$  (vert  $v$ , resp.) dans l'image considérée.  $L$  est le nombre de niveaux utilisés pour quantifier chaque composante couleur, généralement fixé à 256.

Il a été montré que cette version corrigée fournit de meilleurs résultats que ceux fournis par le coefficient initial  $\tau_{RV}$  dans le cadre de la reconnaissance d'objets sous éclairage non contrôlé [81]. Nous utilisons donc ce coefficient pour construire notre descripteur. Cependant, extraire les coefficients  $\tau'_{RV}$ ,  $\tau'_{RB}$  et  $\tau'_{VB}$  de la région considérée ne nous fournit aucune information quant à la distribution spatiale des pixels dans cette région. Nous proposons donc d'étendre ces coefficients (définis pour évaluer les corrélations de rangs entre différentes compo-

santes d'un même ensemble de pixels) à des coefficients qui permettent d'extraire les corrélations de rangs entre les différentes composantes de pixels appartenant à deux ensembles différents. Ainsi, nous proposons de décomposer chaque région détectée en 4 rectangles de tailles identiques sans recouvrement et de calculer les corrélations de rang entre les paires d'ensembles définis par ces rectangles pour chaque paire de composantes. Le paragraphe suivant présente l'extension des coefficients de Kendall aux coefficients qui permettent de calculer des corrélations entre paire d'ensembles.

## 4.4 Corrélation de rangs entre sous-ensembles

Dans cette section, nous définissons la façon d'évaluer la corrélation de rangs entre deux ensembles de pixels, chaque ensemble étant extrait d'un rectangle  $r_i$ ,  $i = 1, \dots, 4$  de la région détectée. De la même manière que précédemment, nous allons illustrer le calcul en travaillant avec les composantes rouge et verte. L'approche peut être bien évidemment étendue à toutes les paires de composantes.

Nous considérons les rectangles  $r_1$  et  $r_2$  extraits de la région détectée et contenant respectivement  $n_{pix_1}$  et  $n_{pix_2}$  pixels. La décomposition particulière retenue (basée sur 4 rectangles de taille identique) conduit à  $n_{pix_1} = n_{pix_2}$ . Au delà de ce cadre particulier, ces rectangles peuvent avoir des tailles différentes. L'objectif est d'évaluer la corrélation entre les mesures de rangs rouges des pixels constituant  $r_1$  et les mesures de rangs verts des pixels constituant  $r_2$ .

L'équation (4.3) nous permet de calculer la corrélation de rangs rouge-vert entre les pixels d'un même ensemble. L'extension à deux sous-ensembles de pixels sans intersection nous contraint de redéfinir le calcul de la valeur  $S$  ainsi que le terme de normalisation qui dépend du nombre de paires totales ainsi que du nombre de paires dont les pixels ont des niveaux (rouges et/ou verts) identiques.

$S$  est définie comme la différence entre le nombre de paires concordantes et le nombre de paires discordantes. Nous allons commencer par expliciter ces nombres

lorsque deux ensembles sont considérés. Dans ce cas, une paire de pixels est constituée d'un pixel  $P1_i$  du rectangle  $r_1$  et d'un pixel  $P2_j$  du rectangle  $r_2$ ,  $i \in [1, n_{pix_1}]$  et  $j \in [1, n_{pix_2}]$ . Elle sera donc notée  $\{P1_i, P2_j\}$ . Les niveaux de rouge de ces pixels seront respectivement notés  $R1_i$  et  $R2_j$  et leurs niveaux de vert  $V1_i$  et  $V2_j$ . Avec ces définitions, l'extension au cas de deux sous-ensembles est immédiate, à savoir si les niveaux de l'intensité rouge et verte des deux pixels  $P1_i$  et  $P2_j$  sont classés dans le même ordre, c'est à dire si  $R1_i < R2_j$  et  $V1_i < V2_j$  ou si  $R1_i > R2_j$  et  $V1_i > V2_j$ , la paire  $\{P1_i, P2_j\}$  est appelée concordante. Sinon, si ces pixels sont de telle sorte que  $R1_i < R2_j$  et  $V1_i > V2_j$  ou que  $R1_i > R2_j$  et  $V1_i < V2_j$ , la paire est appelée discordante. Le nombre total de paires de pixels est égal à  $n_{pix_1} \times n_{pix_2}$ . De la même manière, les termes  $Tk_r$  et  $Tk_v$  de l'équation (4.3) doivent être modifiés et chacun devient égal à la somme des produits des nombres de pixels qui ont un niveau identique dans la composante considérée :

$$Tk_r = \sum_{r=0}^{L-1} t_{1_r} t_{2_r}, \quad (4.4)$$

où  $t_{1_r}$  ( $t_{2_r}$ , resp.) est le nombre des pixels caractérisés par le même niveau rouge  $r$  dans le rectangle  $r_1$  ( $r_2$ , resp.), et

$$Tk_v = \sum_{v=0}^{L-1} t_{1_v} t_{2_v}, \quad (4.5)$$

où  $t_{1_v}$  ( $t_{2_v}$ , resp.) est le nombre de pixels caractérisés par le même niveau vert  $v$  dans la sous-fenêtre  $r_1$  ( $r_2$ , resp.).

Une fois que ces valeurs ont été calculées, le nouveau coefficient de corrélation est défini comme :

$$\tau'_{RV}(r_1, r_2) = \frac{S}{\sqrt{(n_{pix_1} n_{pix_2} - Tk_r)}} \times \frac{1}{\sqrt{(n_{pix_1} n_{pix_2} - Tk_v)}} \quad (4.6)$$

Comme chaque paire de pixels doit être analysée, le calcul de ce coefficient

peut être très gourmand en temps de calculs. Aussi, nous proposons une méthode qui permet d'extraire rapidement ce coefficient des histogrammes 2D couleur de la région. Un histogramme 2D  $H^{k,k'}[r_1]$  du rectangle  $r_1$  est un tableau dont les cellules sont indexées par les niveaux des composantes des pixels. La cellule de  $H^{k,k'}[r_1](u, u')$  indique le nombre de pixels  $P1_i$  dans le rectangle  $r_1$  dont les niveaux  $R1_i$  et  $V1_i$  sont respectivement  $u$  et  $u'$ . Pour l'illustration, nous considérons l'histogramme rouge-vert de  $r_1$  noté  $H^{R,V}[r_1]$  et l'histogramme rouge-vert de  $r_2$  notée  $H^{R,V}[r_2]$ . Ces deux histogrammes sont illustrés dans la figure 4.2.

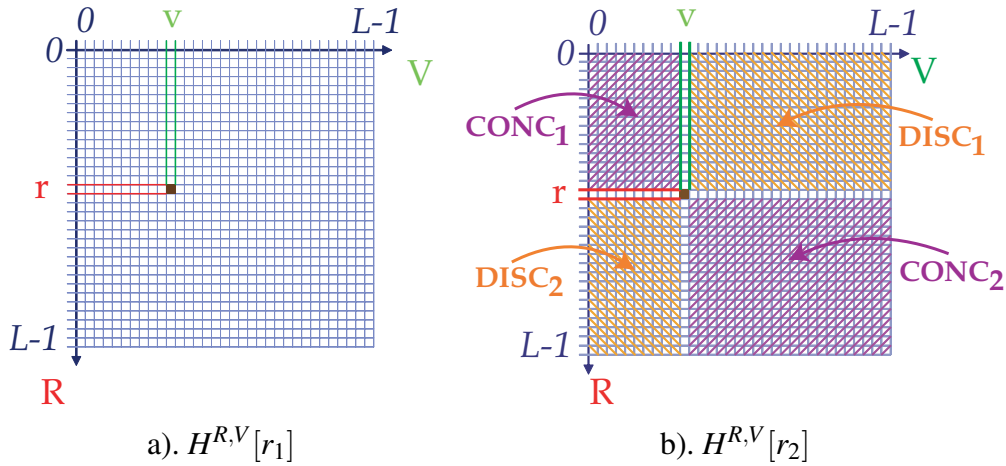


FIGURE 4.2 – Les histogrammes 2D rouge-vert des rectangles  $r_1$  et  $r_2$ .

Nous considérons les pixels caractérisés par le niveau de rouge  $r$  et le niveau de vert  $v$  dans la sous-fenêtre  $r_1$  (cf. figure 4.2 a). A partir des définitions précédentes, nous savons que ces pixels constituent des paires discordantes avec les pixels du rectangle  $r_2$  caractérisés à la fois par des niveaux de rouge inférieurs à  $r$  et des niveaux de vert supérieurs à  $v$  ou caractérisés à la fois par des niveaux de rouge supérieurs à  $r$  et des niveaux de vert inférieurs à  $v$ . Ces pixels de  $r_2$  sont ceux qui sont comptabilisés dans les surfaces  $DISC_1$  et  $DISC_2$  hachurées en orange dans la figure 4.2 b. Par conséquent, le nombre de paires discordantes associées aux niveaux  $(r, v)$  est :

$$\begin{aligned}
 disc(r, v) &= H^{R,V}[r_1](r, v) \times (DISC_1 + DISC_2) \\
 &= H^{R,V}[r_1](r, v) \times \left( \sum_{n_r=0}^{r-1} \sum_{n_v=v+1}^{L-1} H^{R,V}[r_2](n_r, n_v) \right. \\
 &\quad \left. + \sum_{n_r=r+1}^{L-1} \sum_{n_v=0}^{v-1} H^{R,V}[r_2](n_r, n_v) \right). \tag{4.7}
 \end{aligned}$$

De manière à calculer efficacement le nombre de pixels dans les surfaces des histogrammes 2D, nous proposons d'exploiter les histogrammes cumulés 2D. L'idée étant d'évaluer les sommes au préalable comme le font Bay *et al.* [13] avec les images intégrales. L'histogramme cumulé  $HC^{R,V}[r_2]$  est calculé à partir de l'histogramme  $H^{R,V}[r_2]$  comme :

$$HC^{R,V}[r_2](r, v) = \sum_{n_r=0}^r \sum_{n_v=0}^v H^{R,V}[r_2](n_r, n_v) \tag{4.8}$$

pour tout  $\{r, v\} \in [0; L-1]$ .

De cette équation, nous pouvons extraire très rapidement la somme des contenus des cellules des surfaces hachurées dans la figure 4.2 b comme :

$$DISC_1 = HC^{R,V}[r_2](r-1, L-1) - HC^{R,V}[r_2](r-1, v-1), \tag{4.9}$$

et

$$DISC_2 = HC^{R,V}[r_2](L-1, v-1) - HC^{R,V}[r_2](r, v-1). \tag{4.10}$$

Ainsi, l'équation 4.7 devient :

$$\begin{aligned}
 disc(r, v) &= H^{R,V}[r_1](r, v) \times (DISC_1 + DISC_2) \\
 &= H^{R,V}[r_1](r, v) \times (HC^{R,V}[r_2](r-1, L-1) - HC^{R,V}[r_2](r-1, v-1) \\
 &\quad + HC^{R,V}[r_2](L-1, v-1) - HC^{R,V}[r_2](r, v-1)). \tag{4.11}
 \end{aligned}$$



Nous pouvons tenir le même raisonnement pour calculer le nombre de paires concordantes associées aux niveaux  $(r, v)$  :

$$conc(r, v) = H^{R,V}[r_1](r, v) \times (CONC_1 + CONC_2) \quad (4.12)$$

avec

$$CONC_1 = HC^{R,V}[r_2](r-1, g-1) \quad (4.13)$$

et

$$\begin{aligned} CONC_2 = & HC^{R,V}[r_2](L-1, L-1) + HC^{R,V}[r_2](r, v) \\ & - HC^{R,V}[r_2](L-1, v) - HC^{R,V}[r_2](r, L-1) \end{aligned} \quad (4.14)$$

D'où :

$$\begin{aligned} conc(r, v) = & H^{R,V}[r_1](r, v) \times (HC^{R,V}[r_2](r-1, g-1) + HC^{R,V}[r_2](L-1, L-1) \\ & + HC^{R,V}[r_2](r, v) - HC^{R,V}[r_2](L-1, v) - HC^{R,V}[r_2](r, L-1)) \end{aligned} \quad (4.15)$$

Ainsi, pour chaque cellule associée aux niveaux  $(r, v)$  dans l'histogramme  $H^{R,V}[r_1]$ , nous obtenons très vite le nombre de paires concordantes et discordantes et donc la valeur de  $S$  :

$$S = \sum_r \sum_v [conc(r, v) - disc(r, v)]. \quad (4.16)$$

Ainsi, nous venons de proposer une extension du coefficient de corrélation de Kendal qui permet de calculer les corrélations entre composantes couleur de deux ensembles disjoints de pixels. Notons que ce coefficient est théoriquement très peu sensible aux variations d'éclairage alors qu'il est extrait d'histogrammes  $2D$  calculés dans l'espace  $RVB$ , donc très sensibles à ces variations. C'est l'extraction même de ce coefficient qui le rend invariant. Ainsi, aucune transformation vers un espace invariant n'est nécessaire au préalable. Dans le paragraphe suivant, nous allons présenter le descripteur local que nous proposons et qui est fondé sur ce

coefficient.

## 4.5 Le descripteur local fondé sur les corrélations de rang couleur

Comme nous l'avons déjà précisé, nous proposons de diviser la région détectée en 4 rectangles  $r_1, r_2, r_3$  and  $r_4$  de même taille et sans recouvrement. La figure 4.3 montre deux images qui représentent le même objet sous différentes conditions photométriques et radiométriques. Un point d'intérêt (avec échelle et orientation) a été détecté dans chaque image et les rectangles correspondants superposés à l'image. Nous remarquons que lorsque la détection d'un point d'intérêt est correcte, les rectangles correspondants  $r_i$  représentent les mêmes surfaces élémentaires dans les deux images. Ainsi, nous proposons de caractériser la région autour d'un point d'intérêt local en évaluant les corrélations de rangs entre chaque paire de rectangles  $(r_i, r_j)$ ,  $i, j = 1, \dots, 4$  et  $i \neq j$ . Avec 4 rectangles (6 paires) et 3 composantes couleur (3 paires), nous pouvons calculer 18 coefficients de corrélation caractérisant la région détectée :  $\tau'_{RV}(r_1, r_2), \tau'_{RV}(r_1, r_3), \tau'_{RV}(r_1, r_4), \tau'_{RV}(r_2, r_3), \tau'_{RV}(r_2, r_4), \tau'_{RV}(r_3, r_4), \tau'_{RB}(r_1, r_2), \tau'_{RB}(r_1, r_3), \tau'_{RB}(r_1, r_4), \tau'_{RB}(r_2, r_3), \tau'_{RB}(r_2, r_4), \tau'_{RB}(r_3, r_4), \tau'_{VB}(r_1, r_2), \tau'_{VB}(r_1, r_3), \tau'_{VB}(r_1, r_4), \tau'_{VB}(r_2, r_3), \tau'_{VB}(r_2, r_4)$  et  $\tau'_{VB}(r_3, r_4)$ .

Par conséquent, à chaque région d'intérêt détectée, nous associons un vecteur composé de 18 valeurs qui sont peu sensibles aux variations de changement d'éclairage et qui portent une information de répartition des couleurs dans la région détectée.



FIGURE 4.3 – Deux régions d'intérêts correspondantes détectées dans deux images acquises sous différentes conditions d'éclairage.

## 4.6 Conclusion

Dans ce chapitre, nous avons introduit un nouveau descripteur local qui exploite la répartition des couleurs dans un voisinage et qui présente un degré d'invariance relativement élevé puisqu'il suppose qu'un changement des conditions d'acquisition peut être modélisé par une transformation non linéaire des couleurs des pixels. Pour cela, nous avons introduit les corrélations entre les mesures de rangs de pixels appartenant à différents ensembles disjoints. Ces corrélations exploitent les coefficients de corrélation de Kendall qui mesurent la dépendance entre deux variables en analysant leurs mesures de rangs.

Comme les mesures de rangs ne sont que partiellement conservées en cas de changement d'éclairage, nous avons utilisé une mesure de corrélation qui est peu sensible aux légères variations de ces mesures de rangs, rendant ainsi le descripteur final plus stable en cas de métamérisme.

De plus, nous avons proposé une méthode efficace pour extraire rapidement ces coefficients de corrélation à partir des histogrammes  $2D$  de la région détectée et ce à partir de l'espace couleur  $RVB$ . Cette méthode exploite l'idée de Bay *et al.* [13] qui utilisent les images intégrales pour extraire des points d'intérêt.

Dans le cadre de la classification de texture, nous avons développé une approche similaire à partir de matrices de co-occurrences. En effet, nous avons dé-

fini une corrélation de rangs moyenne entre deux composantes pour des paires de pixels situés à une certaine distance et selon une certaine orientation dans l'image. Ainsi, en remplaçant les histogrammes  $2D$  par des matrices de co-occurrences et en adaptant les équations, nous avons mis au point un descripteur de textures couleur qui fournit de bons résultats lorsque les conditions d'acquisition ne sont pas contrôlées [93].

Dans ce chapitre et le précédent, nous sommes partis du principe que les variations photométriques peuvent être modélisées par des transformations linéaires ou non et nous en avons déduit des descripteurs locaux peu sensibles à ces variations. Lorsque ces variations ne peuvent pas être modélisées ou lorsque les couleurs globales des objets peuvent varier, comme dans le cadre de la catégorisation d'objets, où un vélo rouge et un vélo bleu doivent appartenir à une même classe, l'approche consiste plutôt à caractériser le contenu d'une image en décrivant les auto-similarités, c'est à dire en retrouvant les régions qui sont similaires sans en caractériser le contenu. Ceci fait l'objet du chapitre suivant.

---

---

## CHAPITRE 5

---

# L'INVARIANCE PAR AUTO-SIMILARITÉS

### Table des matières

---

<b>5.1</b>	<b>Introduction . . . . .</b>	<b>130</b>
<b>5.2</b>	<b>Les descripteurs fondés sur l'auto-similarité . . . . .</b>	<b>133</b>
5.2.1	Descripteur local d'auto-similarités [95] . . . . .	133
5.2.2	Descripteur global d'auto-similarités [24] . . . . .	136
<b>5.3</b>	<b>Calcul d'auto-similarités invariantes à la rotation . . . . .</b>	<b>140</b>
<b>5.4</b>	<b>Représentation locale de la distribution spatiale des auto-similarités . . . . .</b>	<b>145</b>
<b>5.5</b>	<b>Conclusion . . . . .</b>	<b>146</b>

---

## 5.1 Introduction

Jusqu'à présent, nous sommes partis du principe que les variations des conditions d'acquisition étaient équivalentes pour les couleurs à l'application d'une transformation affine dans le chapitre 3 ou de fonctions croissantes non linéaires dans le chapitre 4. Le cadre de ce chapitre est plus général puisque nous ne posons aucune restriction quant à la transformation qui pourrait modéliser les variations entre deux images dont les contenus sont supposés 'similaires' pour l'application (objets identiques en reconnaissance, objets ou scènes de même classe en classification,...). Les cas de figure suivante justifient par exemple de lever les restrictions sur cette transformation :

- les variations des conditions d'acquisition sont multiples (point de vue, capteur, couleurs des éclairages, directions des éclairages, ombres...).
- les caractéristiques locales comme la couleur ou la texture sont très différentes entre deux contenus 'similaires'. Par exemple, les 3 images 5.1(a), 5.1(h) et 5.1(o) représentent toutes le même symbole mais il n'existe aucune transformation globale entre les couleurs de deux images. Dans ce cas, les descripteurs locaux classiques ne permettent pas de reconnaître ce symbole.
- dans le cadre de la classification d'objets, deux voitures de couleurs différentes et multiples, par exemple, appartiennent à la même classe même s'il n'existe aucune transformation globale entre leurs couleurs.

Dans ces conditions, des travaux récents ont montré qu'il valait mieux décrire le contenu d'une image en représentant localement [19, 95] ou globalement [24] les distributions spatiales des similarités entre régions détectées. C'est ce qu'on appelle les auto-similarités puisque l'on compare les régions d'une même image entre elles et on utilise les répartitions spatiales des mesures de similarité pour décrire cette image. Dans ce chapitre, nous appelons régions les fenêtres carrées

extraites des images, en lien avec les régions d'intérêt exploitées dans les chapitres précédents.

La figure 5.1 présente les répartitions spatiales des auto-similarité de certaines régions extraites des images 5.1(a), 5.1(h) et 5.1(o). Les résultats présentés ici sont ceux fournis par l'approche que nous proposons et que nous allons détailler dans ce chapitre. Ainsi, nous avons extrait 3 régions à des positions correspondantes dans chacune des images 5.1(a), 5.1(h) et 5.1(o). Ces régions sont représentées dans les images 5.1(b), 5.1(d), 5.1(f)}, {5.1(i), 5.1(k), 5.1(m)} et {5.1(p), 5.1(r), 5.1(t)} respectivement. Ensuite, nous avons comparé chacune de ces régions à toutes les régions qui constituent l'image d'origine correspondante. Pour cela, nous avons déplacé une fenêtre glissante de même taille que la région de référence dans l'image correspondante et nous avons à chaque position évalué une mesure de similarité entre la région de référence et le contenu de la fenêtre glissante. Les mesures de similarité obtenues sont représentées dans les images 5.1(c), 5.1(e), 5.1(g)}, {5.1(j), 5.1(l), 5.1(n)} et {5.1(q), 5.1(s), 5.1(u)} respectivement, dans lesquelles une mesure proche de 0 apparaît en noir et une mesure élevée apparaît en blanc. Cet exemple est simplement une illustration qui montre bien que l'utilisation d'auto-similarités permet d'obtenir des représentations similaires pour des images qui ne présentent pas de caractéristiques locales couleur ou de texture similaires.

Deux approches principales ont été récemment proposées pour évaluer des auto-similarités dans une image [24, 95]. La première [95] est locale puisqu'elle compare chaque région détectée aux régions de son voisinage de même taille et retourne la répartition locale des mesures de similarités pour chaque région détectée comme un descripteur local. Dans ce cas, la mesure de similarité entre deux régions est déduite de la sommes des différences des couleurs pixel à pixel. La seconde approche [24] est globale puisque chaque région est comparée à toutes les régions de l'image et le descripteur final représente la distribution spatiale globale des mesures de similarité. Pour accélérer le processus de comparaison, les auteurs

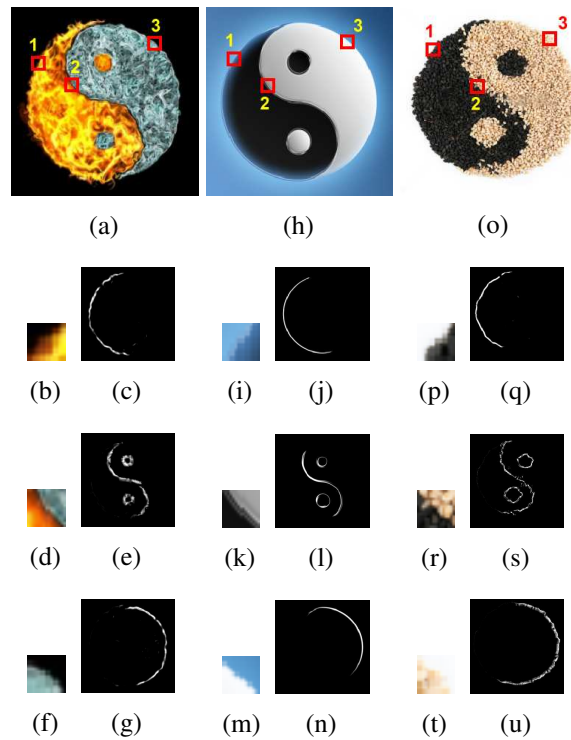


FIGURE 5.1 – Les images 5.1(a), 5.1(h) et 5.1(o) représentent le même symbole mais ne possèdent aucune caractéristique locale de couleur ou de texture similaires. Les régions locales de référence 1 (5.1(b), 5.1(i), 5.1(p)), 2 (5.1(d), 5.1(k), 5.1(r)) et 3 (5.1(f), 5.1(m) et 5.1(t)) sont extraites à partir de positions correspondantes dans les 3 images. Les figures 5.1(c), 5.1(j), 5.1(q), 5.1(e), 5.1(l), 5.1(s), 5.1(g), 5.1(n) et 5.1(u) représentent les similarités (évaluées par notre méthode) entre les régions de référence et leur image respective.



proposent ici d'utiliser l'approche sac-de-mots (cf. annexe A) et d'associer chaque région à un seul mot. Ainsi, la mesure de similarité entre deux régions est binaire, égale à 1 lorsque les mots des deux régions sont les mêmes et 0 lorsqu'ils sont différents. Nous proposons de présenter ces approches dans le détail dans la partie suivante.

Dans ce chapitre, nous proposons une nouvelle méthode pour calculer les auto-similarités dans les images. Pour cela, nous exploitons le descripteur IVC introduit dans le chapitre 3 de ce manuscrit. Cependant, l'originalité de notre approche est de rendre ce descripteur invariant à la rotation de sorte à pouvoir retrouver des auto-similarités entre régions d'une même image qui diffèrent uniquement de par leur orientation. En effet, les deux approches présentées ci-dessus sont très sensibles aux rotations et ne permettent pas de retrouver les auto-similarités comme celles affichées dans les images 5.1(c), 5.1(j), 5.1(q) puisqu'il existe une rotation dans le plan image entre la région de référence et les régions caractérisées par une forte auto-similarité. De plus notre descripteur est très compact et permet donc de comparer très rapidement les régions sans avoir recours à une phase d'apprentissage comme avec les sacs-de-mots et enfin exploite aussi l'information couleur des régions pour les comparer, ce qui permet d'obtenir des mesures d'auto-similarité plus pertinente qu'avec une simple information en niveaux de gris.

## 5.2 Les descripteurs fondés sur l'auto-similarité

### 5.2.1 Descripteur local d'auto-similarités [95]

Le descripteur proposé par Shechtman *et al.* [95] représente la distribution locale des auto-similarités d'une région détectée. Pour cela, les auteurs proposent de comparer chaque région détectée aux régions de même taille distribuées autour de celle-ci dans un voisinage de taille prédéfinie. Ainsi, à chaque pixel de ce voisi-

nage est associée une valeur qui est la mesure de similarité entre la région détectée et la région centrée sur ce pixel. Ceci conduit à une "surface de corrélation" locale comme celle présentée dans la figure 5.2. Cette surface est alors échantillonnée par une grille circulaire caractérisée par 4 distances différentes et 20 angles comme représentée dans cette même figure. Il s'agit d'une représentation en log-polaire qui est ensuite transformée en vecteur de dimension  $4 \times 20 = 80$ . L'échantillonnage consiste à ne retenir que la mesure de similarité la plus élevée dans chaque cellule de la grille log-polaire.

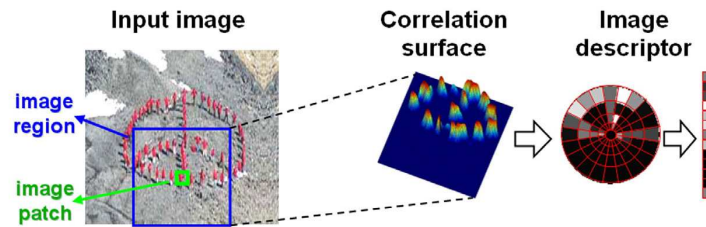


FIGURE 5.2 – Extraction du descripteur local d'auto-similarité. Cette figure a été empruntée à [95]

Dans le détail, voici les étapes qui permettent d'extraire un descripteur local à 80 dimensions :

1. Convertir les composantes *RVB* de l'image couleur en composantes CIE  $L^*a^*b^*$ . Les auteurs ne précisent pas l'illuminant de référence choisi pour cette transformation.
2. Comparer la région détectée notée  $r_P$  (centré sur le pixel  $P$ ) de taille  $5 \times 5$  avec les régions voisines de même taille dans un voisinage circulaire de rayon 40 pixels autour du pixel  $P$ . Pour chaque région voisine  $r_x$ , calculer la somme des carrés des différences  $SSD(r_P, r_x)$  dans l'espace CIE  $L^*a^*b^*$  avec la région de référence  $r_P$ . Ceci fournit la "surface de distances" qui, à chaque position  $x$ , fait correspondre la distance entre  $r_P$  et  $r_x$ .

3. Normaliser la "surface de distance" en "surface de corrélation"  $C_P(x)$  définie comme :

$$C_P(x) = \exp \left( - \frac{SSD(r_P, r_x)}{\max(var_{noise}, var_{auto}(P))} \right), \quad (5.1)$$

où  $var_{noise}$  est une constante fixée par l'utilisateur, elle représente le maximum de variations photométriques acceptables et  $var_{auto}(p)$  est la variance maximale des valeurs de  $SSD$  entre deux régions voisines distantes d'un seul pixel.

4. La surface de corrélation  $C_P(x)$  est ensuite échantillonnée sur une grille log-polaire avec 4 rayons et 20 angles (80 valeurs). Dans chaque cellule, la valeur maximale des valeurs  $C_P(x)$  est retenue.
5. Les 80 valeurs sont normalisées dans l'intervalle  $[0; 1]$  par étirement linéaire afin d'être invariantes aux contenus des régions.

La figure 5.3 présente les grilles log-polaire extraites en 3 positions différentes sur deux images qui représentent le même symbole. Nous pouvons constater que les descripteurs locaux correspondants sont très similaires même si les caractéristiques locales de couleur et de texture des régions concernées sont très différentes d'une image à l'autre.

Cette approche permet d'extraire un ensemble de descripteurs locaux invariants à de larges modifications de textures et de couleurs. Les auteurs proposent ensuite d'appliquer l'approche sac-de-mots à ces descripteurs locaux et à caractériser chaque image par un sac d'auto-similarités locales (BOLSS pour Bag Of Local Self-Similarities).

Cependant, les similarités entre régions d'une même image mais qui ne se situent pas dans un même voisinage (distantes de plus de 40 pixels dans ce cas précis) ne sont pas prises en compte et Deselaers *et al.* proposent donc plutôt d'analyser la répartition spatiale des similarités entre toutes les régions d'une image, sans restriction sur le voisinage local [24].

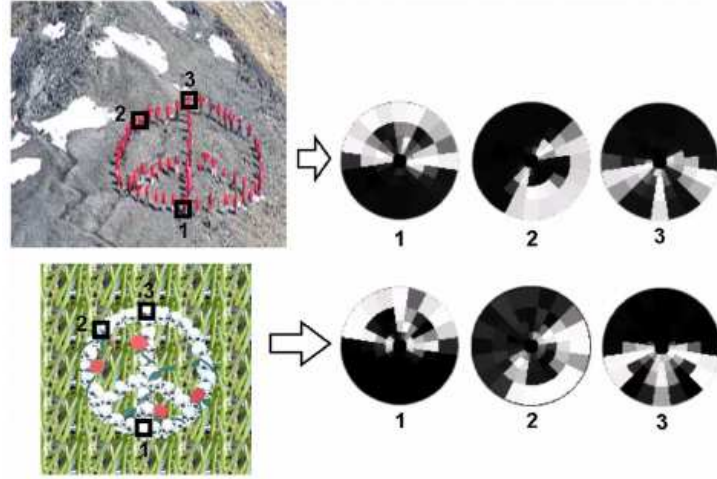


FIGURE 5.3 – Les descripteurs locaux extraits autour de 3 régions correspondantes. Cette figure a été empruntée à [95]

### 5.2.2 Descripteur global d'auto-similarités [24]

Deselaers *et al.* [24] soutiennent l'idée que les auto-similarités doivent être calculées et analysées à l'échelle de l'image et non dans un voisinage local autour de chaque région détectée. Ils proposent donc d'évaluer la similarité entre chaque paire de régions constituant une image et de représenter ces auto-similarités dans des matrices. Concrètement, pour une région  $r_p$  de l'image, il calcule la similarité avec toutes les régions  $r_x$  de l'image et obtient donc une surface de corrélation dont la taille est celle de l'image. Ainsi, pour une image de taille  $H \times W$ , il obtient  $H \times W$  surfaces de corrélation de taille  $H \times W$  chacune. La figure 5.4 représente 2 surfaces de corrélation calculées pour 2 régions correspondantes dans deux images différentes. Une fois encore, nous pouvons remarquer qu'à l'échelle de l'image, les distributions des auto-similarités sont similaires pour ces deux images qui présentent pourtant des caractéristiques locales de couleurs et de textures assez différentes.

Cependant, la prise en compte des auto-similarités à l'échelle de l'image présente deux inconvénients majeurs. D'une part, l'évaluation des similarités entre



FIGURE 5.4 – Répartitions des auto-similarités globales de deux régions dans chaque image. Cette figure a été empruntée à [24]

toutes les paires de régions nécessite un temps de calcul très important et d'autre part, le volume de données nécessaire pour représenter ces distributions globales d'auto-similarités est trop important pour rendre cette approche intéressante dans le cadre de la reconnaissance d'objets.

Deselaers *et al.* ont donc proposé des solutions à ces deux problèmes.

De manière à éviter de comparer les régions pixel à pixel, ils proposent de choisir un descripteur local et de calculer ce descripteur sur un nombre important de régions de la même base ou d'une autre base. Il réalise ensuite une quantification vectorielle dans l'espace des descripteurs pour extraire les mots visuels les plus représentatifs (cf. Annexe A). Ensuite, chaque région est associée au mot visuel le plus proche. La similarité entre deux régions devient alors binaire : 1 lorsque les deux régions sont associées au même mot visuel et 0 sinon. Par conséquent, chaque région est associée à une surface de corrélation binaire.

Pour s'affranchir des problèmes liés au volume de données trop important, les auteurs proposent de sommer les surfaces de corrélation des régions qui appartiennent au même voisinage et de sous-échantillonner ces sommes de surfaces de corrélation.

Dans le détail, voici le processus d'extraction de descripteurs globaux d'auto-

similarités :

1. Calculer les surfaces de corrélation binaires de chaque région. A cette étape, nous avons  $H \times W$  surfaces de corrélation binaires de tailles  $H \times W$ .
2. Projection sur l'image d'une grille de même taille que l'image et constituée de  $D1$  lignes et de  $D2$  colonnes. Les paramètres  $D1$  et  $D2$  sont fixés par l'utilisateur et ne dépendent pas de la taille de l'image considérée. Les surfaces de corrélation des régions qui appartiennent à une même cellule sont sommées pour obtenir  $D1 \times D2$  surfaces de corrélation. Ces surfaces de corrélation sont toujours de taille  $H \times W$  et ne sont plus binaires.
3. Une grille de taille  $H \times W$  est de nouveau projetée sur chacune des  $D1 \times D2$  surfaces de corrélation et la valeur maximale est retenue pour chaque cellule de sorte à obtenir des surfaces de corrélation de taille  $D1 \times D2$ .

Ainsi, à l'issue de ce processus chaque image est associée à  $D1 \times D2$  surfaces de corrélation de taille  $D1 \times D2$ , les paramètres  $D1$  et  $D2$  étant adaptés au besoin de l'application. Ils sont généralement fixés tous les deux à 10. L'une des approches retenues par Deselaers *et al.* consiste à utiliser chaque surface de corrélation de taille  $D1 \times D2$  comme un descripteur "local" de l'image puis à appliquer l'approche sac-de-mots (cf. Annexe A). C'est à dire que les mots visuels sont extraits à partir des descripteurs "surfaces de corrélation" et que chaque image est caractérisée par un sac de surfaces de corrélation (BOCS pour Bag Of Correlation Surfaces).

Ainsi dans les deux approches présentées ici, les similarités entre régions sont évaluées soit en calculant des différences pixel à pixel, soit en comparant les mots visuels associés à chacune des régions. Dans le premier cas, l'évaluation de la mesure de similarité est très gourmande en temps de calcul. Dans le second cas, une étape préalable de création du dictionnaire de mots visuels est nécessaire et la qualité de la base utilisée lors de cette étape est primordiale. De plus, la comparaison de mots visuels retourne une valeur binaire, ce qui peut paraître restrictif dans

certains cas. Enfin, ces deux approches négligent complètement les similarités qui existent entre deux régions qui ont des contenus très similaires à une rotation dans le plan image près. Ce problème est illustré dans la figure 5.5 qui représente une image de vélo avec une région sélectionnée sur le cadre et deux images d'auto-similarités globales calculées avec des méthodes différentes. L'image 5.5(c) représente les auto-similarités retournées lorsque l'on calcule les différences pixel à pixel entre la région sélectionnée et toutes les régions de l'image. Nous observons que seule la partie du cadre orientée de la même manière que la région sélectionnée apparaît dans l'image d'auto-similarités. L'image 5.5(d) représentent les auto-similarités globales retournées par notre méthode qui est invariante à la rotation. Nous voyons qu'une large partie du cadre apparaît dans cette image. Avec l'approche classique (image 5.5(c)), des informations essentielles sont perdues pour reconnaître un vélo alors que notre approche détecte toutes les régions qui sont similaires à la régions sélectionnée, quelle que soit leur orientation. Nous aurions pu illustrer de la manière l'intérêt de l'invariance à la rotation dans les images de la figure 5.1.

Dans la partie suivante, nous proposons de présenter en détail notre méthode de calcul d'auto-similarités. Cette méthode peut être appliquée dans le cas local comme dans le cas global.

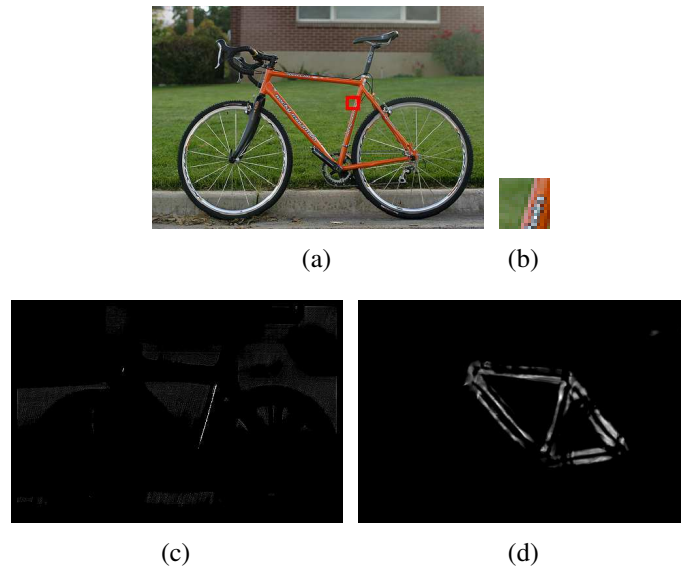


FIGURE 5.5 – Les images 5.5(c) et 5.5(d) représentent les auto-similarités globales entre la région 5.5(b) extraites de l'image 5.5(a) et toutes les régions de cette même image. Les auto-similarités de l'image 5.5(c) sont déduites des différences pixel à pixel entre les régions alors que celles de l'image 5.5(d) sont obtenues grâce à notre approche.

### 5.3 Calcul d'auto-similarités invariantes à la rotation

Lors de l'évaluation d'auto-similarités, plutôt que de comparer les régions elles-mêmes entre elles, nous proposons de comparer des descripteurs de ces régions. L'intérêt des auto-similarités est que ces descripteurs n'ont pas besoin d'être invariants aux conditions d'acquisition puisque nous posons l'hypothèse, comme toutes les méthodes de calculs d'auto-corrélation le font, que les régions d'une même image ont été acquises sous les mêmes conditions. Comme nous savons que tout degré d'invariance superflu peut conduire à une diminution du pouvoir discriminant, le descripteur utilisé doit avoir un degré d'invariance aux variations photométriques presque nul. En revanche, comme le nombre de comparaisons de paires de régions est important, le descripteur utilisé doit être com-



pact. De plus, comme la couleur est une information essentielle dans le cadre de la description locale [17, 98], nous proposons de l'intégrer dans le descripteur. A partir de toutes ces remarques, nous avons décidé d'utiliser le descripteur IVC introduit dans le chapitre 3 de ce manuscrit comme base de description. En effet, ce descripteur peut être très compact, il contient une information couleur et a un degré d'invariance photométrique théorique nul lorsqu'aucune normalisation n'est appliquée.

Nous proposons de rappeler brièvement la construction du descripteur IVC sans normalisation photométrique. Pour une région détectée  $r_1$ , nous construisons deux matrices  $M_{\mathcal{R}_1}$  et  $M_{\mathcal{C}_1}$  qui contiennent respectivement les coordonnées spatiales  $x_i$  et  $y_i$  des pixels  $P_i$  dans la région  $r_1$  et les composantes couleur  $R_i$ ,  $V_i$  et  $B_i$  de ces pixels. Les coordonnées spatiales sont ici les coordonnées relatives à la région analysée, c'est à dire que l'origine est le centre de cette région. Les deux matrices sont définies comme suit :

$$M_{\mathcal{R}_1} = \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ \dots & \dots & \dots \end{bmatrix}, \quad (5.2)$$

et

$$M_{\mathcal{C}_1} = \begin{bmatrix} R_1 & V_1 & B_1 \\ R_2 & V_2 & B_2 \\ \dots & \dots & \dots \end{bmatrix}. \quad (5.3)$$

Ensuite nous déterminons la transformation  $T_{IVC_1}$  de l'espace image vers l'espace couleur telle que :

$$M_{\mathcal{C}_1} \approx M_{\mathcal{R}_1} T_{IVC_1}. \quad (5.4)$$

L'approche des moindres carrés nous fournit la solution suivante :

$$T_{IVC_1} = M_{\mathcal{R}_1}^+ M_{\mathcal{C}_1}, \quad (5.5)$$

où  $M_{\mathcal{R}_1}^+$  est la pseudo-inverse de la matrice  $M_{\mathcal{R}_1}$ .

Le descripteur  $IVC_1$  est obtenu par application de cette transformation aux coordonnées spatiales de trois des coins de la région détectée  $r_1$  :

$$IVC_1 = \begin{bmatrix} -\frac{sx}{2} & -\frac{sy}{2} & 1 \\ -\frac{sx}{2} & \frac{sy}{2} & 1 \\ \frac{sx}{2} & \frac{sy}{2} & 1 \end{bmatrix} T_{IVC_1}, \quad (5.6)$$

où  $sx$  et  $sy$  sont les dimensions de la région  $r_1$ .

Cependant, dans le cadre du calcul d'auto-similarités, les régions sont des fenêtres carrées non orientées (à la différence des régions extraites par des détecteurs comme Harris-Laplace ou Hessien-Laplace). Par conséquent, le descripteur IVC ainsi extrait n'est pas invariant aux rotations appliquées dans le plan image. En effet, si nous appliquons au préalable à la région  $r_1$  une rotation  $2D$   $Rot$  dans le plan image, la transformant ainsi en une région  $r_2$  telle que :

$$M_{\mathcal{R}_2} = M_{\mathcal{R}_1} Rot, \quad (5.7)$$

et

$$M_{\mathcal{C}_2} = M_{\mathcal{C}_1}. \quad (5.8)$$

Nous obtenons la transformation  $T_{IVC_2}$  par l'équation :

$$\begin{aligned} T_{IVC_2} &= M_{\mathcal{R}_2}^+ M_{\mathcal{C}_2}, \\ T_{IVC_2} &= (M_{\mathcal{R}_1} Rot)^+ M_{\mathcal{C}_1}, \\ T_{IVC_2} &= Rot^T M_{\mathcal{R}_1}^+ M_{\mathcal{C}_1}, \\ T_{IVC_2} &= Rot^T T_{IVC_1}, \end{aligned} \quad (5.9)$$

car la matrice  $Rot$  étant une matrice de rotation, nous avons  $Rot^T Rot = Identit$ , elle peut donc être sortie de la pseudo-inverse.

Ainsi, sans connaissance a priori sur l'orientation de la région, nous considé-

rons que les coordonnées des 3 coins de  $r_2$  sont  $(-\frac{sx}{2}, -\frac{sy}{2})$ ,  $(-\frac{sx}{2}, \frac{sy}{2})$  et  $(\frac{sx}{2}, \frac{sy}{2})$  et nous obtenons un descripteur  $IVC_2$  différent de  $IVC_1$ .

Pour rendre notre descripteur invariant à la rotation, nous proposons de décomposer chaque transformation  $T_{IVC_k}$  en un ensemble de transformations "élémentaires" (rotation, translation, cisaillement, ...) et d'annuler la partie correspondant à la rotation. Pour cela, nous ré-écrivons l'équation (5.4) pour un seul pixel en décomposant la transformation  $T_{IVC_k}$  :

$$\begin{aligned}
 [R_1 V_1 B_1] &\approx [x_1 y_1 1] T_{IVC_k} \\
 [R_1 V_1 B_1] &\approx [x_1 y_1 1] \begin{bmatrix} mk_1 & mk_2 & mk_3 \\ mk_4 & mk_5 & mk_6 \\ tk_1 & tk_2 & tk_3 \end{bmatrix} \\
 [R_1 V_1 B_1] &\approx [x_1 y_1] \begin{bmatrix} mk_1 & mk_2 & mk_3 \\ mk_4 & mk_5 & mk_6 \end{bmatrix} + \begin{bmatrix} tk_1 & tk_2 & tk_3 \end{bmatrix} \\
 [R_1 V_1 B_1] &\approx [x_1 y_1] Mk_{2 \times 3} + Tk_{1 \times 3},
 \end{aligned} \tag{5.10}$$

où les  $mk_i$  et les  $tk_j$  sont les paramètres de la transformation  $T_{IVC_k}$ .

Cette décomposition permet d'isoler la translation  $Tk_{1 \times 3}$  des autres transformations élémentaires  $Mk_{2 \times 3}$  qui constituent la transformation  $T_{IVC_k}$ . De l'équation (5.9), nous pouvons déduire que  $M2_{2 \times 3} = Rot^T M1_{2 \times 3}$ . L'objectif est donc maintenant d'extraire la rotation des matrices  $M1_{2 \times 3}$  et  $M2_{2 \times 3}$ . Pour cela, nous proposons d'utiliser la méthode de factorisation QR qui permet de décomposer une matrice  $Mk_{2 \times 3}$  en un produit d'une matrice de rotation  $Qk_{2 \times 2}$  et une matrice triangulaire supérieure droite  $Rk_{2 \times 3}$  telles que  $Mk_{2 \times 3} = Qk_{2 \times 2}Rk_{2 \times 3}$ . Ainsi, nous décomposons les deux matrices  $M1_{2 \times 3}$  et  $M2_{2 \times 3}$  avec cette méthode QR telle que :

$$M1_{2 \times 3} = Q1_{2 \times 2}R1_{2 \times 3}, \tag{5.11}$$

et

$$M_{2 \times 3} = Q_{2 \times 2} R_{2 \times 3}, \quad (5.12)$$

et nous en déduisons que :

$$\begin{cases} Q_{2 \times 2} = Rot^T Q_{1 \times 2}, \\ R_{2 \times 3} = R_{1 \times 3}. \end{cases} \quad (5.13)$$

En effet, de par la définition de la décomposition QR, la rotation de la matrice  $M_{2 \times 3}$  est isolée dans la matrice  $Q_{2 \times 2}$ . Par conséquent, la rotation  $2D$   $Rot$  qui existe dans le plan image entre les deux régions  $r_1$  et  $r_2$  n'affecte que les matrices  $Q_{1 \times 2}$  et  $Q_{2 \times 2}$  et laisse invariantes les matrices  $R_{1 \times 3}$  et  $R_{2 \times 3}$ . L'objectif étant d'extraire les informations invariantes à toute rotation dans le plan image, nous proposons de constituer le descripteur d'auto-similarité  $DAS_k$  des 3 valeurs de la matrice  $T_{k \times 3}$  (éq. (5.10)) et des 5 valeurs non nulles parmi les 6 de la matrice  $R_{k \times 3}$ . Ce descripteur est donc composé de 8 valeurs invariantes à la rotation. Pour des raisons d'homogénéité, chaque dimension est divisée par son écart-type calculé sur des descripteurs évalués sur un nombre important de régions extraites d'images diverses.

Pour résumer, le processus d'évaluation du descripteur  $DAS$  d'une région  $r$  est le suivant :

- évaluer la transformation  $T_{VC}$  de l'espace image vers l'espace couleur (cf. chapitre 3),
- décomposer cette transformation en deux transformations  $M_{2 \times 3}$  (rotation, échelle, cisaillement) et  $T_{1 \times 3}$  (translation) (éq. (5.10)),
- appliquer la décomposition QR à la matrice  $M_{2 \times 3}$  pour obtenir les deux matrices  $Q_{2 \times 2}$  et  $R_{2 \times 3}$ ,
- conserver les 3 valeurs de  $T_{1 \times 3}$  et les 5 valeurs non nulles de  $R_{2 \times 3}$  pour constituer le descripteur  $DAS$ ,

- normaliser chaque dimension par son écart-type calculé sur un nombre important de descripteur.

Ainsi, les avantages de notre descripteur sont triples. Premièrement comme il représente à la fois les couleurs et leur distribution spatiale relative dans l'espace image, il est très discriminant et peut donc déterminer si deux régions sont similaires ou non. Deuxièmement, le temps de traitement nécessaire pour évaluer la similarité entre deux régions est très faible étant donné que chaque descripteur est constitué de seulement 8 valeurs. Troisièmement, ce descripteur est totalement invariant à la rotation dans l'espace image. Le pouvoir discriminant et la propriété d'invariance de rotation de notre descripteur peuvent être vérifiés dans les figures 5.1 et 5.5 puisque les auto-similarités élevées sont bien obtenues presque uniquement pour des paires de régions qui présentent des répartitions spatiales de couleurs similaires à une rotation près.

Cette méthode d'extraction d'auto-similarités dans une image peut être utilisée pour des descripteurs globaux ou locaux tels que ceux présentés dans le paragraphe 5.2. Dans la partie suivante, nous proposons un descripteur local spécifiquement défini pour notre mesure d'auto-similarités.

### 5.4 Représentation locale de la distribution spatiale des auto-similarités

L'idée est ici de représenter pour chaque région  $r_0$  (typiquement de taille  $5 \times 5$ ) d'une image donnée la distribution spatiale des auto-similarités dans son voisinage (typiquement de taille  $80 \times 80$ ). Considérons une région  $r_1$  de même taille dans ce voisinage, nous retenons 3 valeurs pour représenter les positions relatives de ces régions : 2 valeurs  $\Delta x_{0-1}$  et  $\Delta y_{0-1}$  représentent les translations suivant les axes  $x$  et  $y$ , respectivement, et 1 valeur  $\Delta \theta_{0-1}$  représente l'angle de rotation entre ces deux régions. Les valeurs de translation sont facilement obtenues en évaluant les

différences entre les coordonnées  $x$  et  $y$  des centres des régions. Le  $\Delta\theta$  peut être obtenu à partir des matrices  $Q0_{2 \times 2}$  et  $Q1_{2 \times 2}$  présentées dans la section précédente. En effet, nous avons vu que s'il existe une rotation  $Rot$  entre les deux régions dans l'espace image alors nous avons  $Q2_{2 \times 2} = Rot^T Q1_{2 \times 2}$ . Ceci nous permet d'écrire :

$$Rot = Q1_{2 \times 2} Q2_{2 \times 2}^T, \quad (5.14)$$

car ce sont des matrices rotation (leurs transposées sont donc égales à leurs inverses).

La matrice de rotation  $Rot$  ainsi obtenue nous permet de déterminer l'angle  $\Delta\theta_{0-1}$  qui existe entre les deux régions puisque cette matrice contient les valeurs de cosinus et sinus de cet angle.

Ainsi, nous proposons de créer une structure  $3D$  (appelé volume d'auto-similarité) associée à la région  $r_0$  dont les axes sont  $\Delta x$ ,  $\Delta y$  et  $\Delta\theta$  et d'insérer la mesure de similarité entre les régions  $r_0$  et  $r_1$  dans la cellule de coordonnées  $(\Delta x_{0-1}, \Delta y_{0-1}, \Delta\theta_{0-1})$ . De la même manière, nous pouvons insérer les mesures de similarité entre  $r_0$  et toutes les régions  $r_i$  de son voisinage afin de représenter la répartition spatiales des auto-similarités autour de  $r_0$ . Typiquement, nous discrétisons les  $\Delta x$  et  $\Delta y$  sur 5 intervalles et les  $\Delta\theta$  sur 4 intervalles pour obtenir un volume d'auto-similarités de taille  $5 \times 5 \times 4 = 100$  associé à chaque région  $r_0$ . Suivant la méthode de Shechtman *et al.* [95], nous conservons la mesure de similarité maximale dans chaque cellule du volume.

## 5.5 Conclusion

Dans ce chapitre, nous avons présenté une nouvelle méthode qui permet de caractériser la distribution spatiale des auto-similarités dans une image. Dans un premier temps, nous avons proposé d'associer un descripteur compact à chaque région afin d'extraire rapidement les similarités entre régions. Ce descripteur est

## CHAPITRE 5. L'INVARIANCE PAR AUTO-SIMILARITÉS

---

fondé sur l'approche exposée dans le chapitre 3 de ce manuscrit et permet de représenter la distribution locale des couleurs tout en étant invariant à la rotation dans l'espace image. Dans un second temps, nous avons proposé une structure à 3 dimensions pour décrire la distribution spatiale de ces auto-similarités. Cette structure informe sur la translation et la rotation qui existe entre deux régions similaires.





---

---

## CHAPITRE 6

---

# RÉSULTATS EXPÉRIMENTAUX

### Table des matières

---

<b>6.1</b>	<b>Introduction . . . . .</b>	<b>149</b>
<b>6.2</b>	<b>Les bases d'images . . . . .</b>	<b>149</b>
6.2.1	La base d'images pour l'appariement . . . . .	149
6.2.2	Les bases d'images pour la classification . . . . .	151
<b>6.3</b>	<b>Les descripteurs testés . . . . .</b>	<b>159</b>
6.3.1	Propriétés théoriques d'invariance . . . . .	159
6.3.2	Compacités . . . . .	160
<b>6.4</b>	<b>Appariement de régions locales . . . . .</b>	<b>161</b>
6.4.1	Paramètres de IVC et CVI . . . . .	162
6.4.2	Comparaison avec les descripteurs de la littérature . . .	169
<b>6.5</b>	<b>Classification d'objets . . . . .</b>	<b>181</b>
<b>6.6</b>	<b>Conclusion . . . . .</b>	<b>186</b>

---

## 6.1 Introduction

L'objectif de ce chapitre est d'évaluer les performances de nos descripteurs sur des tests expérimentaux et de les comparer avec les descripteurs locaux de la littérature. Afin de tester les descripteurs dans différents contextes, nous proposons de mener deux expériences différentes. La première consiste à tester le pouvoir discriminant et le degré d'invariance des descripteurs dans le cadre de la comparaison de régions locales extraites d'images acquises sous des conditions différentes. La seconde teste les descripteurs dans le contexte de la classification d'objets qui est fondée sur l'approche sacs de mots (cf. Annexe A).

Ainsi, nous allons, dans la seconde partie, présenter les bases d'images utilisées pour mener ces tests. Dans la troisième partie, nous présenterons les descripteurs que nous avons choisis de comparer avec les nôtres. Les deux parties suivantes seront alors consacrées aux résultats des deux expériences. Nous finirons par une conclusion sur ce chapitre.

## 6.2 Les bases d'images

### 6.2.1 La base d'images pour l'appariement

Pour l'appariement de régions, la base d'images d'Amsterdam ALOI est utilisée [43]. Cette base contient 1000 objets dans des images acquises sous des conditions différentes. Chaque objet est acquis sous toutes les conditions proposées : variations de la couleur de l'éclairage, de la direction de l'éclairage, du point de vue, du taux de compression Jpeg et du flou appliqué aux images. Les images de la figure 6.1 présentent 2 objets sous les différentes conditions.

Van de Sande *et al.* avaient déjà souligné dans [98] que les variations photométriques de la base ALOI ne perturbent pas beaucoup les couleurs puisque les descripteurs qui ne présentent aucune invariance théorique fournissent de bons

résultats de reconnaissance. Ils ont donc proposé de créer des perturbations artificielles sur les couleurs pour tester les propriétés d'invariance de leurs descripteurs. Pour ces tests, nous proposons aussi d'appliquer des perturbations contrôlées aux couleurs des images de la base ALOI afin d'accroître les difficultés. De plus, cette approche permet de vérifier les propriétés d'invariance théoriques des descripteurs puisque les transformations appliquées dans l'espace couleur sont connues. Ainsi, nous proposons d'appliquer aux couleurs des images une transformation affine telle que celle-ci :

$$\begin{bmatrix} R' \\ V' \\ B' \end{bmatrix} = \begin{bmatrix} a_1 & b_1 & b_2 \\ b_3 & a_2 & b_4 \\ b_5 & b_6 & a_3 \end{bmatrix} \cdot \begin{bmatrix} R \\ V \\ B \end{bmatrix} + \begin{bmatrix} o_1 \\ o_2 \\ o_3 \end{bmatrix}, \quad (6.1)$$

pour laquelle les valeurs des paramètres  $a_i$ ,  $b_j$  et  $o_k$  vont être contrôlées.

A partir de ce modèle de transformation et en faisant varier les valeurs des paramètres, nous proposons 4 nouveaux tests, chacun caractérisé par l'application d'une transformation dans l'espace couleur : transformation diagonale, translation, transformation diagonale + translation, transformation affine. Pour chaque test, nous allons avoir 7 conditions de variation  $C_v$ ,  $v = 1, \dots, 7$ , dont la difficulté va croître avec la valeur de  $v$ . Pour ces tests, la condition de référence est celle caractérisée par un illuminant de couleur 2175 K et les images de test sont les images acquises sous un illuminant légèrement différent de couleur 2250 K auxquelles nous avons appliqué ces transformations dans l'espace couleur. Les valeurs des paramètres  $a_i$ ,  $b_j$  et  $o_k$  sont précisées dans le tableau 6.1. Lorsque les valeurs sont présentées comme des intervalles dans ce tableau, cela signifie que nous avons choisi aléatoirement des valeurs dans cet intervalle. Les images de la figure 6.2 présentent l'effet de ces variations artificielles sur les couleurs de deux objets.

Nous devons noter que les transformations couleur appliquées artificiellement aux images ne modifient pas de manière choquante le contenu de ces images (cf figure 6.2). Les variations que nous avons appliquées restent dans les mêmes in-

## CHAPITRE 6. RÉSULTATS EXPÉRIMENTAUX

		$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$	$C_7$
Diagonale	$a_i$	[0.95 1.05]	[0.85 0.95]	[0.75 0.85]	[0.62 0.72]	[0.45 0.55]	[0.35 0.45]	[0.28 0.38]
	$b_j$	0	0	0	0	0	0	0
	$o_k$	0	0	0	0	0	0	0
Translation	$a_i$	1	1	1	1	1	1	1
	$b_j$	0	0	0	0	0	0	0
	$o_k$	[-2 2]	[-6 -2]	[-10 -6]	[-14 -10]	[-18 -14]	[-22 -18]	[-27 -23]
Diagonale+Translation	$a_i$	[0.95 1.05]	[0.85 0.95]	[0.75 0.85]	[0.62 0.72]	[0.45 0.55]	[0.35 0.45]	[0.28 0.38]
	$b_j$	0	0	0	0	0	0	0
	$o_k$	[-2 2]	[-6 -2]	[-10 -6]	[-14 -10]	[-18 -14]	[-22 -18]	[-27 -23]
Affine	$a_i$	1	1	1	1	1	1	1
	$b_j$	[0 -0.07]	[-0.07 -0.14]	[-0.14 -0.21]	[-0.21 -0.29]	[-0.29 -0.36]	[-0.36 -0.43]	[-0.43 -0.5]
	$o_k$	[-2 2]	[-6 -2]	[-10 -6]	[-14 -10]	[-18 -14]	[-22 -18]	[-27 -23]

TABLE 6.1 – Valeurs des paramètres de la transformation affine de l'équation (6.1) pour les conditions  $C_v$ ,  $v = 1, \dots, 7$  des 4 tests.

tervalles que ceux proposés par van de Sande *et al.* [98]. Cette variation artificielle est aussi motivée par le fait que la base ALOI ne permet pas de tester l'invariance des descripteurs face aux changements de systèmes d'acquisition. Or nous savons qu'une modification des sensibilités spectrales des capteurs a un impact important sur les couleurs qui pourrait être modélisé par l'application d'une transformation affine puisqu'il est similaire à une modification des primaires lorsque l'on passe d'un espace colorimétrique à un autre. Van de Weijer *et al.* [100] soutiennent d'ailleurs l'idée que dans ce contexte, les descripteurs doivent être invariants aux rotations dans l'espace des couleurs : "*For the design of physically meaningful color features not only the invariance with respect to spatial coordinate changes is desired but also the invariance with respect to color coordinate systems rotations. Features based on different measurement devices which measure the same spectral space should yield the same results*".

### 6.2.2 Les bases d'images pour la classification

Dans le contexte de la classification d'objets, nous avons utilisé 4 bases différentes qui sont Flowers [5], Soccers [7], Birds [3] et Pascal Voc 2010 [6].

### **Flowers**

Cette base d'images contient 17 catégories de fleurs avec 80 images par catégorie (cf. Figure 6.3). Pour cette base les informations de couleur et de forme sont importantes toutes les deux pour différencier les classes de fleurs. Pour nos tests, nous avons choisi 40 images par classe pour l'apprentissage et 40 images pour le test.

### **Birds**

Cette base d'images contient 6 catégories d'oiseaux avec 100 images par catégorie (cf. Figure 6.4). Pour cette base les informations de couleur et de forme sont importantes pour différencier les classes mais la forme est prédominante. Pour nos tests, nous avons choisi 60 images par classe pour l'apprentissage et 40 images pour le test.

### **Soccers**

Cette base d'images contient 7 catégories de joueurs de football avec 40 images par catégorie (cf. Figure 6.5). Pour cette base l'information de couleur est prédominante pour différencier les classes. Pour nos tests, nous avons choisi 25 images par classe pour l'apprentissage et 15 images pour le test.

### **Pascal VOC 2010**

Cette base d'images contient 20 catégories avec des nombres variables d'images par catégorie (cf. Figure 6.6). Pour cette base l'information de couleur n'apporte pas beaucoup d'information supplémentaire lorsqu'elle est ajoutée à l'information de forme pour différencier les classes. Pour nos tests, nous avons choisi 3072 images au total pour l'apprentissage et 3084 images pour le test.

## CHAPITRE 6. RÉSULTATS EXPÉRIMENTAUX

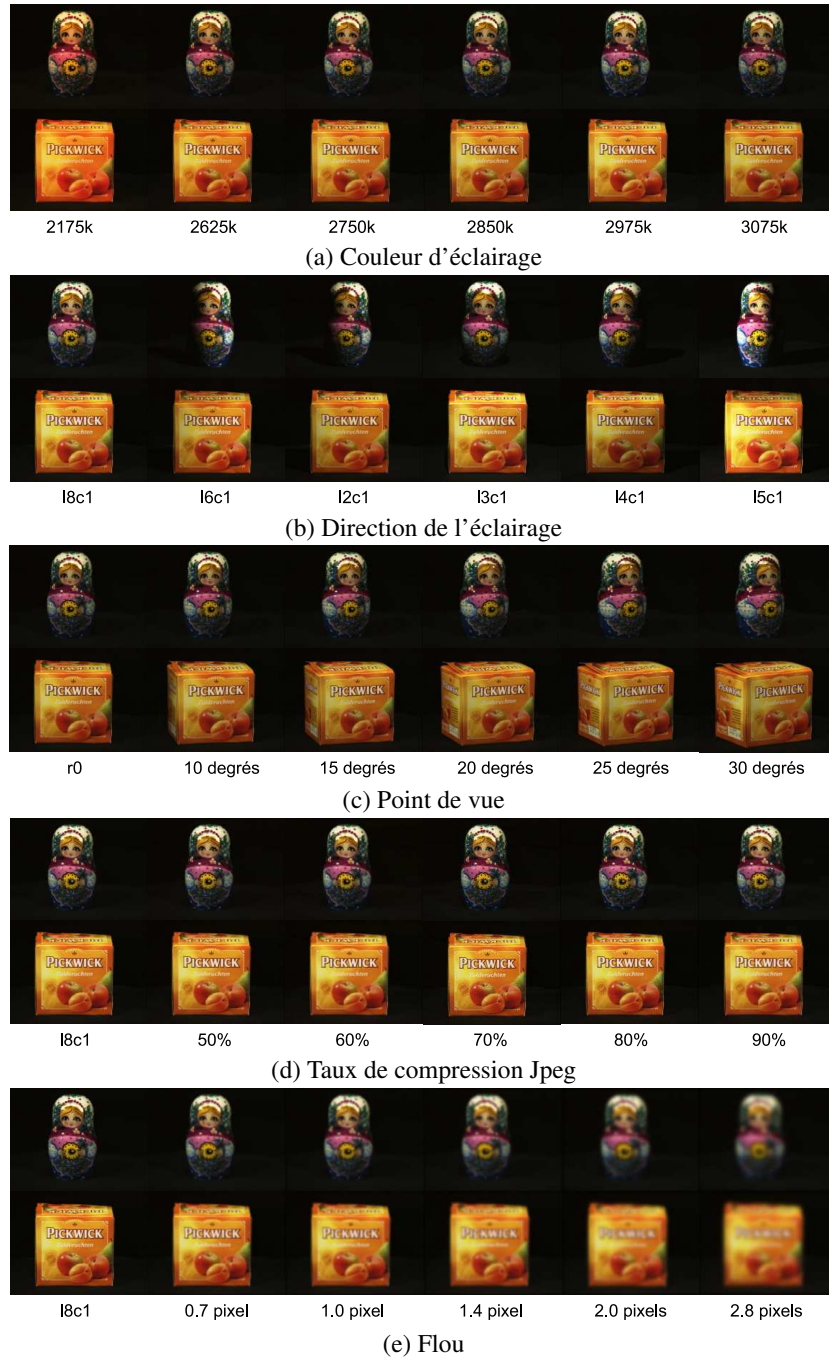


FIGURE 6.1 – Deux objets de la base ALOI sous différentes conditions.

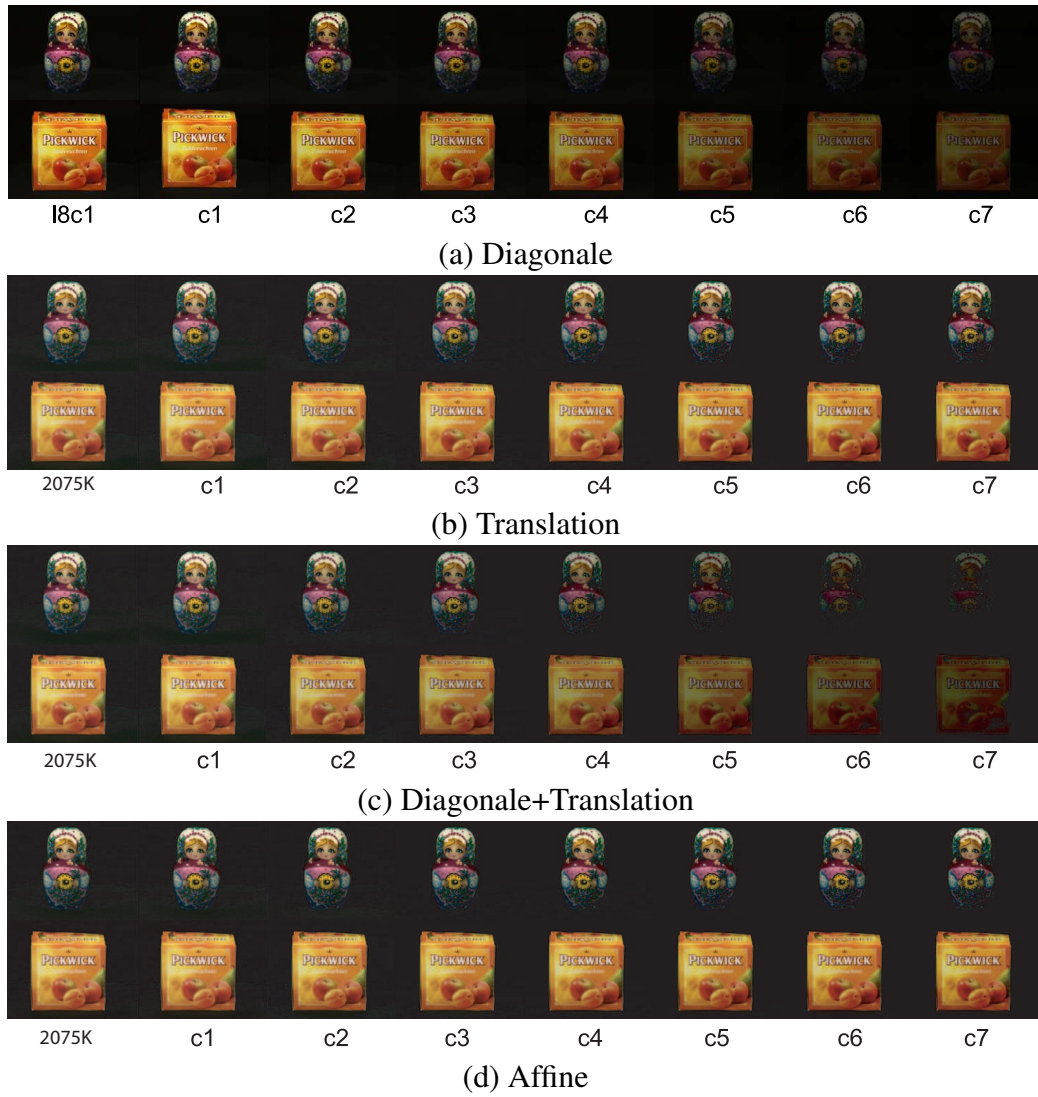


FIGURE 6.2 – Les conséquences des variations artificielles sur les couleurs de deux images.





FIGURE 6.3 – Quelques images de la base Flowers





FIGURE 6.4 – Quelques images de la base Birds



## CHAPITRE 6. RÉSULTATS EXPÉRIMENTAUX



FIGURE 6.5 – Quelques images de la base Soccers



FIGURE 6.6 – Quelques images de la base VOC 2010

## 6.3 Les descripteurs testés

### 6.3.1 Propriétés théoriques d'invariance

De manière à comparer les performances de nos descripteurs avec l'état de l'art, nous avons sélectionné les 6 descripteurs locaux qui fournissent les meilleurs résultats parmi plus de 25 descripteurs locaux testés dans [17, 98, 101] :

- *SIFT* niveaux de gris [68] avec une invariance théorique à l'intensité de l'éclairage et aux translations dans l'espace couleur,

4 *SIFT* couleur :

- *HSV-SIFT* [15] sans invariance théorique,
- *Opponent-SIFT* [98] avec une invariance théorique à l'intensité de l'éclairage et aux translations dans l'espace couleur,
- *C-color-SIFT* [17] avec une invariance théorique à l'intensité de l'éclairage et aux contours d'ombre,
- *RGB-SIFT* avec une invariance théorique à la couleur de l'éclairage,

un descripteur issu de la concaténation entre *SIFT* et l'histogramme de teinte :

- *Hue-SIFT* [101] avec une invariance théorique à l'intensité de l'éclairage et aux translations dans l'espace couleur.

Pour les *SIFT* couleur, nous avons utilisé le code fournit par Mikolajczyk [8] et pour le *Hue-SIFT* celui de van de Weijer [9].

Nous rappelons que notre descripteur *CORR* a une invariance théorique à toute application de fonctions croissantes sur les composantes couleur. Le descripteur *CVI* est invariant à toute transformation affine appliquée dans l'espace couleur. Le degré d'invariance du descripteur *IVC* peut être contrôlé par le nombre de degrés de liberté de la fonction de normalisation associée. Le lecteur se reportera au chapitre 3 pour plus de détails.

### 6.3.2 Compacités

Les résultats des descripteurs dans l'expérience d'appariement ainsi que dans l'expérience de classification doivent être analysés en conservant à l'esprit les dimensions de chaque descripteur. Sur la figure 6.2, nous pouvons remarquer que les descripteurs proposés sont très compacts en comparaison avec les autres descripteurs testés. Cette compacité peut aider (et est parfois requise) dans de nombreuses applications comme le suivi d'objets dans les vidéos ou la classification d'objets qui nécessite une étape de quantification vectorielle dans l'espace des descripteurs pour créer les dictionnaires visuels. De par nos expériences, nous avons pu constater que la création d'un dictionnaire visuel fondé sur les SIFT couleur demande un temps de calcul très important.

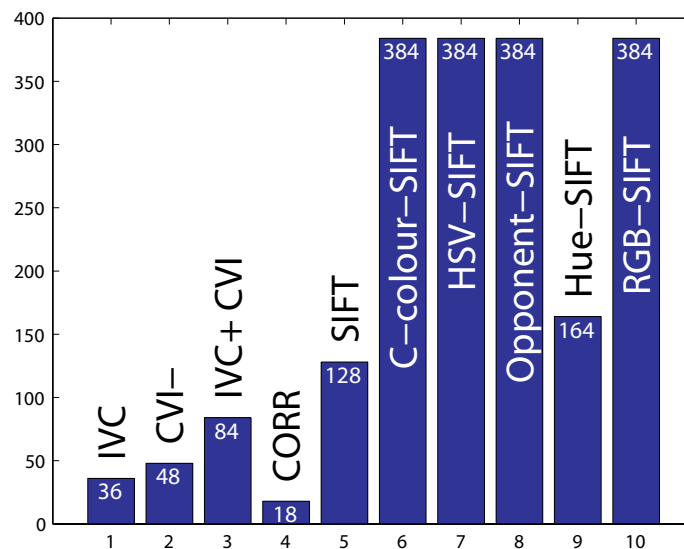


TABLE 6.2 – Dimensions des descripteurs testés



## 6.4 Appariement de régions locales

Dans cette expérience, nous utilisons le même protocole que celui proposé par Burghout *et al.* [17]. Pour cela, nous utilisons la base ALOI qui est présentée plus haut.

Des régions ont été extraites des images par le détecteur de Harris-Affine [72] et Burghout *et al.* ont sélectionné les plus stables (une par image) en terme de positions, échelles et orientations à travers les différentes variations [4]. L'objectif étant que les régions extraites des images du même objet représentent les mêmes éléments de surface de cet objet. Ainsi, nous obtenons 1000 régions, chacune d'elle étant acquise sous toutes les conditions proposées par la base ALOI. Le test consiste à comparer les images de ces régions sous différentes conditions d'acquisition afin de mesurer le pouvoir discriminant et les propriétés d'invariance de chaque descripteur testé.

Le protocole et les critères d'évaluation sont similaires à ceux utilisés par Burghout *et al.* [17] et van de Sande *et al.* [98]. Tout d'abord, l'appariement par le plus proche voisin est adopté en utilisant la distance Euclidienne. Cependant, pour le descripteur obtenu par concaténation de IVC et CVI, chaque dimension est divisée par son écart-type évalué une seule fois sur environ 12000 régions. Pour les conditions de variation de la base ALOI (couleur de l'éclairage, direction de l'éclairage, point de vue, compression Jpeg et flou), nous considérons 6 ensembles constitués d'un nombre identique de régions. Pour les variations artificielles, nous avons 8 ensembles. Un ensemble est appelé l'ensemble de référence  $Ens_R$  et les 5 (ou 7) autres sont les ensembles de tests  $Ens_{Ti}$ ,  $i = 1, \dots, 5$  ou  $i = 1, \dots, 7$ . Chaque ensemble contient des régions extraites d'images acquises sous des conditions identiques. Les ensembles de tests sont constitués de telle sorte que les différences de conditions entre l'ensemble de référence  $Ens_R$  et l'ensemble de test  $Ens_{Ti}$  augmentent lorsque  $i$  augmente. Pour chaque ensemble de test considéré indépendamment, nous comparons chaque région de référence avec toutes les régions de cet

ensemble de test et nous évaluons le pourcentage d'appariements corrects (PAC), c'est à dire le pourcentage de régions de référence dont la distance avec la région de test correspondante (celle qui représente les mêmes éléments de surface) est la minimum parmi les distances avec toutes les autres régions. Ces PAC sont équivalents à la notion de "rappel" dans ce contexte où chaque région de référence à une seule région correspondante dans chaque ensemble de test. Nous aurions pu aussi analyser les courbes de rappel-précision mais nous avons remarquer que celles-ci n'apportent pas d'informations supplémentaires, donc nous avons préféré nous concentrer sur les courbes qui représentent l'évolution du PAC en fonction des difficultés croissantes, c'est à dire que nous traçons le PAC obtenu pour chaque ensemble de test, en faisant croître  $i$ . L'allure de ces courbes nous permet d'avoir une idée à la fois sur le pouvoir discriminant du descripteur testé et sur ses propriétés d'invariance à la condition considérée. En effet, les PAC obtenus pour de faibles variations des conditions (les premiers points à gauche sur la courbe) renseignent sur le pouvoir discriminant du descripteur alors que l'évolution de la courbe vers des conditions de plus en plus difficiles (les points à droite) informe sur les propriétés d'invariance de ce descripteur. Le tableau 6.3 résume ces informations.

### 6.4.1 Paramètres de IVC et CVI

Dans le chapitre 3 nous avons introduit les deux descripteurs IVC et CVI et nous avons pu constater que certains paramètres de ces descripteurs pouvaient être réglés. Dans cette partie, nous proposons de mesurer l'impact de ces paramètres sur leur pouvoir discriminant et leur sensibilité aux variations des conditions d'acquisition. Pour cela, nous allons comparer les performances obtenues sur la base ALOI en fonction de différentes valeurs de ces paramètres.

Les paramètres testés sont les suivants :

- la taille de CVI, liée à la décomposition de la région considérée ( $1 \times 1$ ,  $2 \times 2$ ,  $3 \times 3$ ,  $4 \times 4$  ou  $5 \times 5$  dans la figure 6.7),

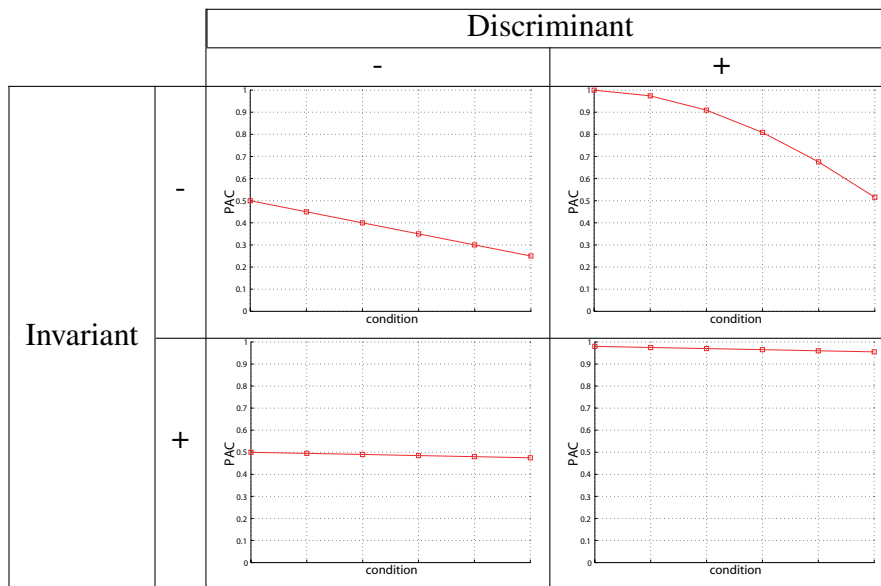


TABLE 6.3 – Evolution des courbes de PAC en fonction de l’invariance et du pouvoir discriminant du descripteur testé.

- la taille de IVC, liée à la décomposition de la région considérée ( $1 \times 1$ ,  $2 \times 2$  ou  $4 \times 4$  dans la figure 6.8),
- le nombre de degrés de liberté de la transformation utilisée pour normaliser IVC (0ddl, 3ddl, 6ddl ou 12ddl dans les figures 6.9 et 6.10),
- la dimension de l’espace couleur ( $1 \times 3D$  pour *RVB* ou  $3 \times 2D$  pour *RV*, *RB* et *VB* dans la figure 6.11),
- l’utilisation d’une contrainte géométrique ou non pour l’évaluation de IVC en  $3 \times 2D$  (avec contrainte ou sans contrainte dans la figure 6.11).

### La taille de CVI

Dans la figure 6.7 nous avons représenté les résultats obtenus sous diverse conditions. Ces résultats doivent être analysés relativement à la taille des descripteurs testés de manière à s’assurer que toute augmentation de la taille soit accompagnée d’un accroissement conséquent des résultats. Ainsi, les tailles des



descripteurs testés sont les suivantes :

- $\text{CVI-}1 \times 1 : 3 - D$ ,
- $\text{CVI-}2 \times 2 : 12 - D$ ,
- $\text{CVI-}3 \times 3 : 27 - D$ ,
- $\text{CVI-}4 \times 4 : 48 - D$ ,
- $\text{CVI-}5 \times 5 : 75 - D$ .

Les courbes nous montrent que, en moyenne, mis à part  $\text{CVI-}1 \times 1$ , les descripteurs fournissent de bons résultats qui varient peu entre les descripteurs. Les différences apparaissent davantage pour les variations de la direction de l'éclairage. La spécificité de ces variations est qu'elles font apparaître des ombres et des reflets dans certaines régions, conduisant à un éclairage non homogène au sein d'une même région. Nous avons montré que le descripteur CVI est invariant aux transformations affines appliquées dans l'espace couleur à toutes les couleurs de la région analysée. En effet, si la transformation affine est appliquée à une seule partie des couleurs de la région, c'est à dire si la variation photométrique n'est pas uniforme sur l'ensemble des pixels de la régions, l'invariance ne sera pas conservée. Ainsi, lorsque nous décomposons une région en sous-régions, la contrainte d'uniformité de la variation photométrique est plus souple puisqu'elle est limitée aux sous-régions. Autrement dit, moins l'éclairage est homogène spatialement, plus il faut réduire la taille des régions sur lesquelles est calculé le CVI. C'est ce que nous pouvons observer sur la courbe représentant les variations de direction de l'éclairage. Cependant, dans ce cas précis, les résultats obtenus par la décomposition  $4 \times 4$  sont presque équivalents à ceux de la décomposition  $5 \times 5$ , le passage de  $48 - D$  à  $75 - D$  ne semble donc pas justifié. En conclusion, après analyse de tous ces résultats, nous proposons d'opter pour le descripteur  $\text{CVI-}4 \times 4$ . C'est celui-ci qui sera utilisé dans la suite des expériences. Nous avons repéré en bleu les courbes de ce descripteur dans la figure 6.7.

### La taille d'IVC

La figure 6.8 présente les résultats obtenus pour différentes tailles du descripteur IVC :

- IVC- $1 \times 1$  :  $9 - D$ ,
- IVC- $2 \times 2$  :  $36 - D$ ,
- IVC- $4 \times 4$  :  $144 - D$ .

Les courbes présentent des résultats relatifs similaires à ceux de la section précédente. Nous constatons aussi que le passage de IVC- $2 \times 2$  à IVC- $4 \times 4$  ne permet pas d'augmenter de manière conséquente les résultats et nous optons donc pour le descripteur IVC- $2 \times 2$  pour la suite des expériences. Nous avons repéré en bleu les courbes de ce descripteur dans la figure 6.8.

### La normalisation d'IVC

Les figures 6.9, 6.10 présentent les performances obtenues par le descripteur IVC- $2 \times 2$  en fonction du nombre de degrés de liberté de la fonction de normalisation. Nous avons vu dans le chapitre 3 que ce nombre de degrés de liberté va directement influencer le degré d'invariance du descripteur. Dans le cadre de cette étude, nous avons donc complétés les tests réalisés précédemment sous les 5 conditions de la base ALOI, par les 4 conditions que nous avons artificiellement créées. En effet, l'objectif de ces tests est de mesurer l'impact d'une transformation appliquée dans l'espace couleur sur les performances des descripteurs.

Nous rappelons les modèles de variations photométriques auxquels se rapportent les différents nombres de degrés de liberté :

- 0 dll : aucune invariance théorique,
- 3 dll : modèle diagonal (équ. (2.20)),
- 6 dll : modèle diagonal et translation (équ. (2.21)),

- 12 dll : modèle affine (éq. (2.23)).

Pour les 5 courbes de la figure 6.9, nous constatons l'importance de la prise en compte du compromis entre invariance et pouvoir discriminant. En effet, sur ces courbes, il apparaît que plus nous conférons un degré d'invariance élevé à un descripteur, moins celui est discriminant. Et même pour les courbes qui concernent les variations photométriques, comme les changements de couleur ou de direction de l'éclairage, le degré d'invariance photométrique le plus faible est requis puisque le descripteur IVC-0 dll fournit les meilleurs résultats. Cela dit, concernant la couleur de l'éclairage, la compression et le flou, les résultats fournis par IVC-0 dll, IVC-3 dll et IVC-6 dll sont similaires. En moyenne, cependant, sur les 5 conditions testées, le meilleur choix sera celui du descripteur IVC-0 dll, donc celui qui ne présente aucune invariance photométrique théorique.

Comme les variations photométriques de la base ALOI ne perturbent pas beaucoup les couleurs, nous proposons de poursuivre ce test sur les images auxquelles nous avons appliqué artificiellement des transformations affines dans l'espace couleur comme l'ont fait van de Sande *et al.* [98]. Les résultats de ces tests sont affichés dans la figure 6.10.

Dans la figure 6.10, nous constatons que la théorie est validée. En effet, le descripteur IVC-0 dll montre très peu d'invariance aux transformations appliquées, mis à part à la translation. Le descripteur IVC-3 dll qui présente une invariance théorique uniquement aux transformations diagonales voit ses résultats chuter dès lors qu'une translation est appliquée. Le descripteur IVC-6 dll présente les meilleurs résultats en moyenne sur l'ensemble des conditions testées. Il obtient en effet des résultats presque parfaits dans toutes les conditions sauf lorsque la transformation affine (avec termes non diagonaux non nuls) est appliquée. Quant au descripteur IVC-12 dll, son invariance n'est pas remise en cause par les tests mais son pouvoir discriminant n'est pas aussi élevé que celui de IVC-6 dll. Ceci est dû à son degré d'invariance très élevé ce qui ne lui permet pas de respecter le compromis invariance/pouvoir discriminant. Nous constatons d'ailleurs

que ses performances sont stables quelles que soient les conditions de test, mais reste à un niveau moyen inférieur à IVC-6 dll. Ainsi, le descripteur IVC-6 dll est celui qui gère le mieux le compromis entre pouvoir discriminant et degré d'invariance. Nous retenons ce descripteur lorsque les conditions d'acquisition créent des variations photométriques importantes. Lorsque ces variations sont faibles (cf. figure 6.9), le descripteur IVC-0 dll présente les meilleures performances.

### **La dimension de l'espace couleur**

Il nous a paru opportun de comparer les résultats obtenus par l'application d'une transformation de l'espace image vers un espace couleur 3D (IVC-3D) à ceux obtenus par les applications de 3 transformations de l'espace image vers un espace couleur 2D *RV*, *RB* et *VB* (IVC-2D). L'idée étant que le passage de la 2D à la 3D pouvait être perturbé par l'apparition d'une dimension supplémentaire. Initialement, nous avons d'ailleurs travaillé uniquement avec des espaces 2D. Cela dit, en intégrant les 3 dimensions dans une seule transformation, nous prenons en compte l'aspect vectoriel de la couleur, ce qui intuitivement pourrait accroître le pouvoir discriminant. Dans la figure 6.11, nous comparons les résultats des descripteurs IVC-3D-0ddl et IVC-2D-sans-contrainte. nous devons noter que les dimensions de ces descripteurs sont différents. IVC-3D-0ddl est un descripteur plus compact puisqu'il possède 36 valeurs et IVC-2D-sans-contrainte est un descripteur à 72 valeurs. Nous remarquons que ces deux descripteurs fournissent les mêmes résultats sur la bases ALOI, ce qui valide l'hypothèse que la prise en compte de la couleur comme une information 3D permet d'accroître le pouvoir discriminant puisque 36 valeurs permettent d'obtenir les mêmes résultats que 72 pour le descripteur qui considère 3 plans 2D.

### **L'utilisation d'une contrainte géométrique**

L'intérêt de travailler dans 3 plans 2D réside dans le fait que nous pouvons imposer une contrainte géométrique à la transformation de l'espace image vers

l'espace couleur (cf. section 3.3.3 du chapitre 3). L'intérêt de cette contrainte étant d'annuler une invariance géométrique inutile du descripteur afin d'accroître son pouvoir discriminant. En effet, de par la détection de régions d'intérêt dans les images, la transformation qui existe entre deux régions qui représentent les mêmes éléments de surface peut être modélisée par une rotation, un changement d'échelle et une translation, mais la partie cisaillement de la transformation affine 2D modélisée par notre transformation  $T_{IVC}$  n'a pas lieu d'être. Nous avons donc souhaité annuler cette invariance et tester l'intérêt de cette annulation. Nous comparons donc les résultats obtenus par les descripteurs IVC-2D-sans-contrainte et IVC-2D-avec-contrainte dans la figure 6.11. Sur les 5 conditions testées, nous constatons que l'ajout de la contrainte n'a un impact que lorsque le point de vue varie entre les images. c'est effectivement à ce niveau qu'était attendu l'amélioration puisqu'elle traite des déformations géométriques. Cependant l'impact de cette contrainte étant peu conséquent, et étant donné que nous ne pouvons pas placer une telle contrainte lorsque l'on considère l'espace couleur 3D, nous proposons de retenir le descripteur IVC-3D-oddll dans les tests de la figure 6.11. Les performances de ce descripteur apparaissent en bleu sur les différentes courbes de cette figure.

Dans cette partie, nous avons fait varier les paramètres de nos descripteurs couleur afin d'optimiser le choix des valeurs de ces paramètres. En conclusion, pour le descripteur CVI nous retenons celui calculé sur une décomposition  $4 \times 4$  de la région détectée. Sa dimension sera donc de 48. Pour le descripteur IVC, de nombreux paramètres ont été testés et nous choisissons le descripteur calculé sur une décomposition  $2 \times 2$  de la région, en travaillant avec un espace couleur 3D. Sa dimension sera donc de 36. Pour la normalisation, lorsque les variations photométriques sont faibles, le descripteur IVC-0 dll sera choisis, sinon ce sera le descripteur IVC-6 dll.

Nous avons conduit ces expériences sur une base d'image de référence contenant 1000 objets dans des images acquises sous des conditions différentes. Ces

choix peuvent donc être retenus comme les choix par défaut. Cependant, la souplesse de nos descripteurs permet de les adapter à d'autres bases. Nous avons fait en sorte de présenter l'impact de chaque paramètre sur les performances des descripteurs de manière à ce que le lecteur puisse faire un choix en connaissance de cause en fonction des priorités des critères de qualité (pouvoir discriminant, degré d'invariance, compacité).

### 6.4.2 Comparaison avec les descripteurs de la littérature

Dans ce paragraphe, nous proposons de comparer les descripteurs retenus dans la partie précédente aux descripteurs locaux les plus performants parus dans la littérature [17, 98]. Nous testons les performances de tous ces descripteurs sur les 5 conditions de variation fournies par la base ALOI ainsi que sur les transformations artificielles que nous avons proposées. Les résultats sont regroupés dans les figures 6.12 et 6.13. Dans le but de simplifier l'analyse, nous avons ajouté sur chaque graphique un diagramme à bâtons qui représente pour chaque descripteur la moyennes de bons résultats sur la condition testée.

Tout d'abord, nous soulignons que nous obtenons des résultats d'appariement très similaires à ceux présentés par les articles récents [17, 98] pour les descripteurs de la littérature. L'unique différence est que nous obtenons des résultats légèrement supérieurs pour le Hue-SIFT que [98]. Comme les 100 et 500 régions sont choisies de manière aléatoire, nous ne pouvions pas obtenir exactement les mêmes résultats.

Pour aider à l'analyse, nous rappelons que, mis à part pour les tests de changement de point de vue, les régions détectées dans les images contiennent exactement les mêmes éléments de surface au cours des variations d'une condition. En effet, comme l'objet n'a pas bougé entre les acquisitions sous différentes conditions, la détection de régions a été faite dans les images de référence, sous une condition et les mêmes positions, orientations et échelles ont été extraites dans les images acquises sous d'autres conditions. Ainsi, les erreurs de détection n'inter-

viennent que dans le cas des variations du point de vue, pour lequel les positions des éléments de surface varient.

En moyenne, les résultats relativement décevants de notre descripteur CORR sous certaines conditions s'expliquent par le fait que les mesures de rang sont très instables dans plusieurs cas. Par exemple, lorsque l'éclairage n'est pas homogène dans une région (c'est le cas lorsque la direction de l'éclairage varie), les niveaux de composantes couleur de certains pixels varient différemment de ceux d'autres pixels, perturbant complètement les mesures de rang de tous les pixels de la région. Lorsqu'il y a des erreurs de détections (c'est le cas pour les variations de point de vue), l'apparition de nouveaux éléments de surface et/ou la disparition de certains autres ont aussi tendance à modifier les mesures de rang des pixels. Les variations comme la compression jpeg ou le flou ont tendance à homogénéiser les couleurs des pixels localement et perturbent ainsi les mesures de rang. Enfin, une transformation affine dont les coefficients peuvent être négatifs ne peut pas toujours être assimilée à une fonction croissante qui est pourtant l'hypothèse nécessaire à la préservation des mesures de rang. Ainsi, le descripteur CORR est plutôt recommandé lorsque la couleur de l'éclairage n'est pas contrôlé et lorsque les transformations dans l'espace des couleurs sont des translations et/ou des applications de matrices diagonales. Il faut noter que ce descripteur à 18 dimensions est le plus compact de tous et qu'il obtient sur certaines courbes des résultats similaires à des descripteurs de dimensions 384.

Concernant les descripteurs CVI et IVC, nous notons que les résultats valident la théorie. En effet, nous remarquons que IVC fournit de meilleurs résultats que CVI lorsque les transformations sont géométriques (points de vue, Jpeg et flou) alors que CVI fournit des résultats similaires ou meilleurs que IVC dans le cas de transformations photométriques (couleur de l'éclairage, direction de l'éclairage et application de transformations dans l'espace couleur). Ceci vérifie la théorie puisque nous avons montré que IVC est invariant aux transformations dans l'espace image alors que CVI est invariant aux transformations dans l'espace couleur.

De plus, ce qui est intéressant de noter est que la simple concaténation des deux descripteurs fournit des résultats qui sont toujours équivalents à ceux obtenus par le meilleur des deux descripteurs. Cette concaténation qui reste compact (84-D) semble donc être la meilleure solution lorsque les conditions de variations ne sont pas connues. En effet, nous remarquons que ce descripteur (IVC+CVI) fournit en moyenne les meilleurs résultats de tous les descripteurs testés.

En ce qui concerne la couleur de l'éclairage, tous les descripteurs fournissent de très bons résultats avec une légère supériorité de la concaténation IVC+CVI. Comme van de Sande *et al.* l'avaient déjà signalé [98], l'invariance théorique aux variations de la couleur de l'éclairage n'est pas requise pour qu'un descripteur fournisse de bons résultats sur cette base. En effet, les descripteurs tels que IVC-Oddl ou les Opponent-SIFT ne sont pas invariants mais fournissent de très bons résultats dans la pratique. De plus, nous pouvons noter que l'addition de la couleur aux SIFT niveaux de gris n'apporte pas beaucoup sous cette condition puisque les SIFT couleur ont des résultats comparables aux SIFT niveaux de gris.

Les variations de direction d'éclairage sont celles qui perturbent le plus les descripteurs testés. Comme nous l'avons déjà précisé, ces modifications conduisent à l'apparition d'ombres, de reflets et de non-homogénéité de l'éclairage au sein des régions. Nous voyons que le descripteur CVI supporte le mieux ces variations. Ceci vient du fait que ce descripteur, d'une part, est complètement invariant à toute transformation affine dans l'espace couleur, et d'autre part, il décompose chaque région en  $4 \times 4$  sous-régions et caractérise indépendamment chacune de ces 16 sous-régions. Ainsi, de par le calcul de ce descripteur, les 16 sous-régions sont indépendamment normalisées par rapport aux variations des conditions d'acquisition. Les descripteurs qui utilisent SIFT décomposent aussi la région en  $4 \times 4$ , mais la normalisation est effectuée sur la région globale et non sur chaque sous-région. Ceci explique pourquoi le descripteur CVI offre les meilleures performances dans ces conditions. Une fois encore, nous notons que la concaténation IVC+CVI exploite les avantages du meilleur des deux descripteurs CVI et IVC



puisqu'elle fournit les meilleurs résultats de tous les descripteurs testés.

Comme nous l'avons déjà précisé, dans le cas de variations de point de vue, il peut y avoir quelques erreurs de positions, d'échelles et d'orientations des ellipses détectées, même si nous avons considéré les ellipses les plus stables (fournies par Burghout *et al.*) au cours des variations. Par conséquent les descripteurs doivent prendre en compte les légères erreurs de détection ainsi que des transformations affines dans l'espace 3D. Comme le descripteur SIFT niveaux de gris est fondé sur les orientations de gradients locaux, il est plus sensible à des variations de point de vue et aux conséquences d'une mauvaise détection. C'est pour cela que l'intégration d'information couleur est nécessaire pour augmenter les performances de ce descripteur. Ainsi, la plupart des SIFT couleur fournissent de meilleurs résultats que le SIFT niveau de gris. Cependant, les meilleurs résultats sont fournis par le descripteur IVC et ceci permet de confirmer l'invariance théorique de ce descripteur à toute transformation affine appliquée dans l'espace image. D'après les résultats, nous remarquons que le compromis entre pouvoir discriminant et invariance n'est pas correctement géré par le descripteur CVI en cas de variations de point de vue. En effet, comme pratiquement aucune invariance photométrique n'est requise pour ces variations, l'invariance très forte de CVI conduit à de moins bonnes performances, équivalentes à celles du SIFT niveaux de gris. Heureusement, ces résultats ne dégradent pas de manière conséquente les performances de la concaténation IVC+CVI puisque ce descripteur fournit les meilleurs résultats après IVC en comparaison avec tous les autres descripteurs testés.

La compression Jpeg et le flou causent des distortions géométriques dans l'espace image qui ne peuvent pas être modélisées par des transformations affines. C'est pour cela que le descripteur CVI présente de moins bons résultats que les autres descripteurs testés. De plus, nous savons que la compression Jpeg détériore davantage les composantes chromatiques des pixels que leur composante intensité. L'ajout d'information couleur n'est donc pas forcément pertinente dans ce cas. Les très bonnes performances du SIFT niveaux de gris illustrent cela. Ces

deux raisons expliquent pourquoi la concaténation IVC+CVI ne fournit pas les meilleurs résultats en cas de compression Jpeg. Heureusement, la plupart des images sont rarement compressées à des taux aussi importants que ceux testés ici. Pour des taux faibles, nous remarquons que les résultats de la concaténation IVC+CVI sont moins perturbés par les mauvaises performances de CVI.

Sur la figure 6.13, les variations ont été artificiellement créées dans les images en appliquant des transformations dans l'espace couleur. Nous remarquons que le descripteur CVI est complètement insensible à ces variations puisqu'il obtient les résultats maximum à chaque fois. Pour le descripteur IVC, nous avons ici utilisé une normalisation à 6 degrés de liberté comme nous l'avons noté dans les tests préalables sur ce descripteur. Nous remarquons que la concaténation IVC+CVI présente aussi presque toujours des performances maximales (100%) avec une très légère baisse en cas de transformation affine.

## CHAPITRE 6. RÉSULTATS EXPÉRIMENTAUX

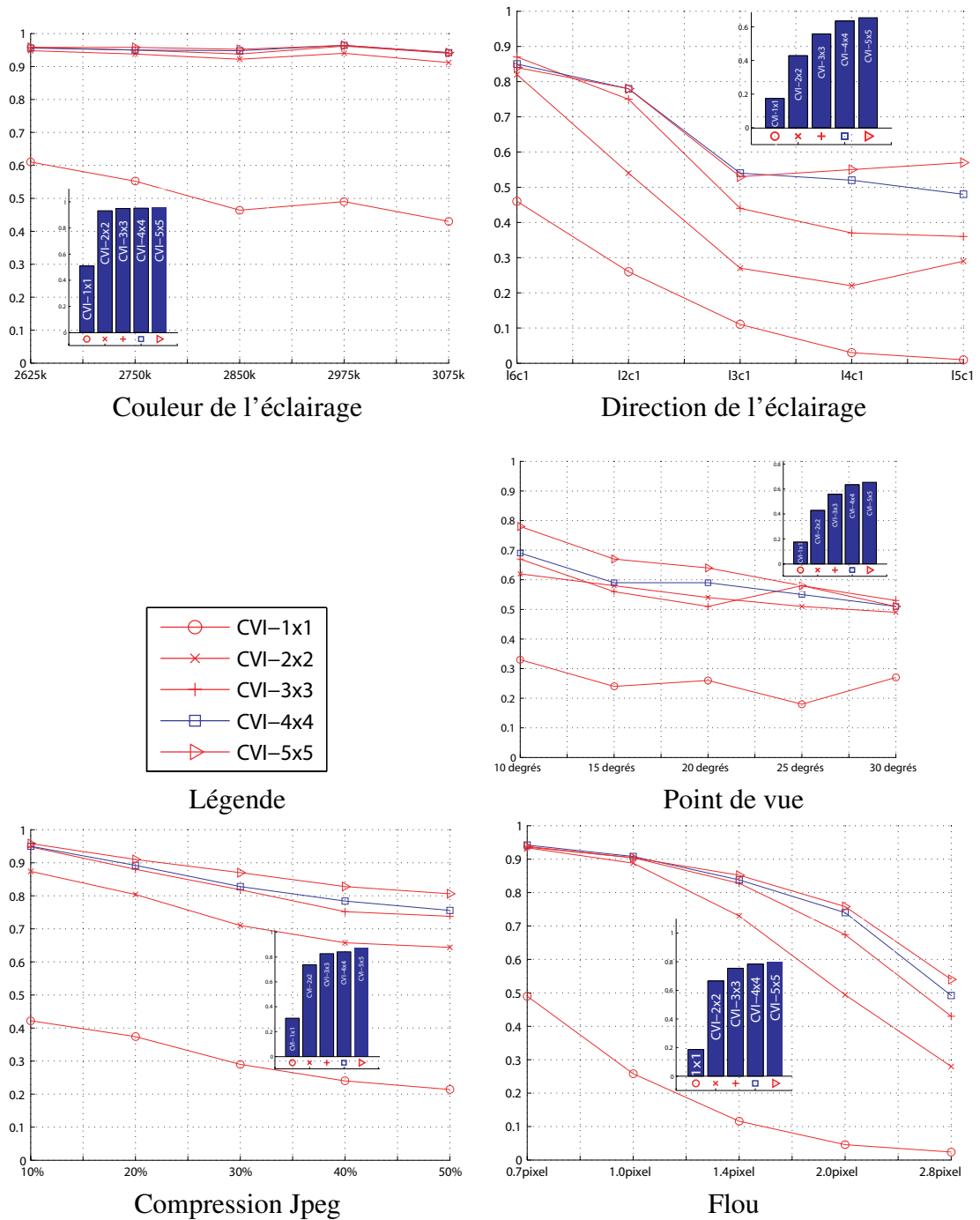


FIGURE 6.7 – La taille de CVI sur la base ALOI

## CHAPITRE 6. RÉSULTATS EXPÉRIMENTAUX

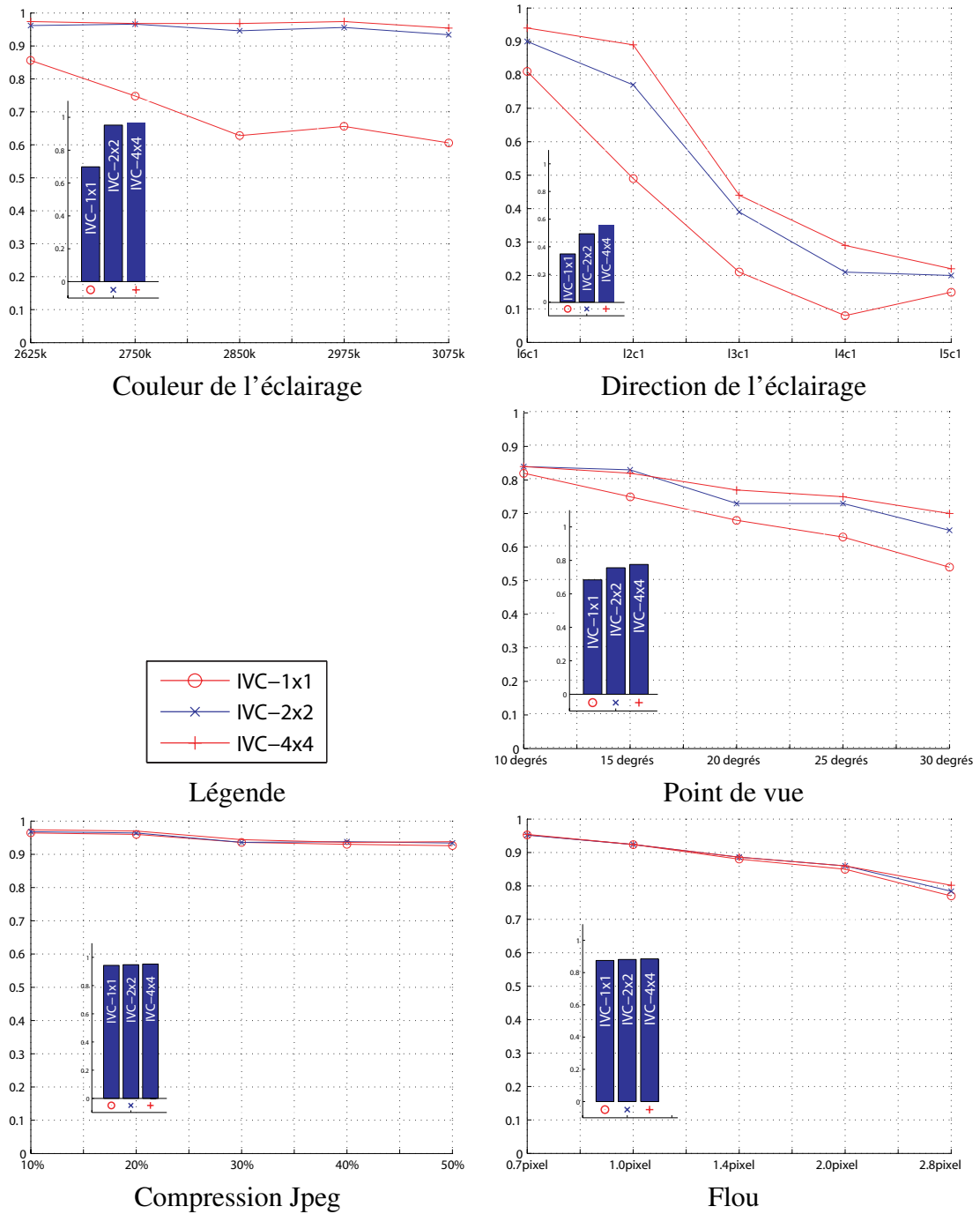


FIGURE 6.8 – La taille d'IVC sur la base ALOI

## CHAPITRE 6. RÉSULTATS EXPÉRIMENTAUX

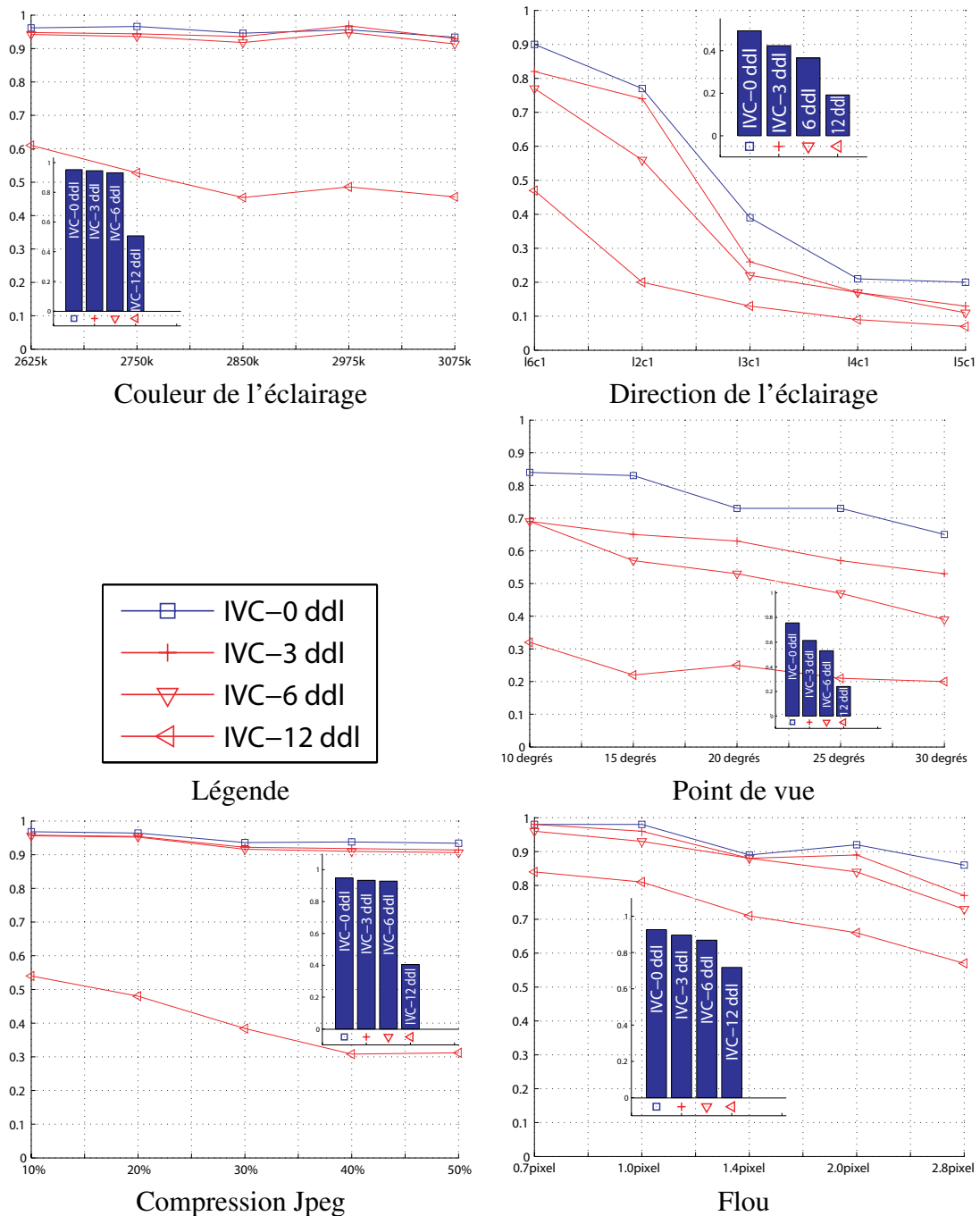


FIGURE 6.9 – La normalisation d'IVC sur la base ALOI

## CHAPITRE 6. RÉSULTATS EXPÉRIMENTAUX

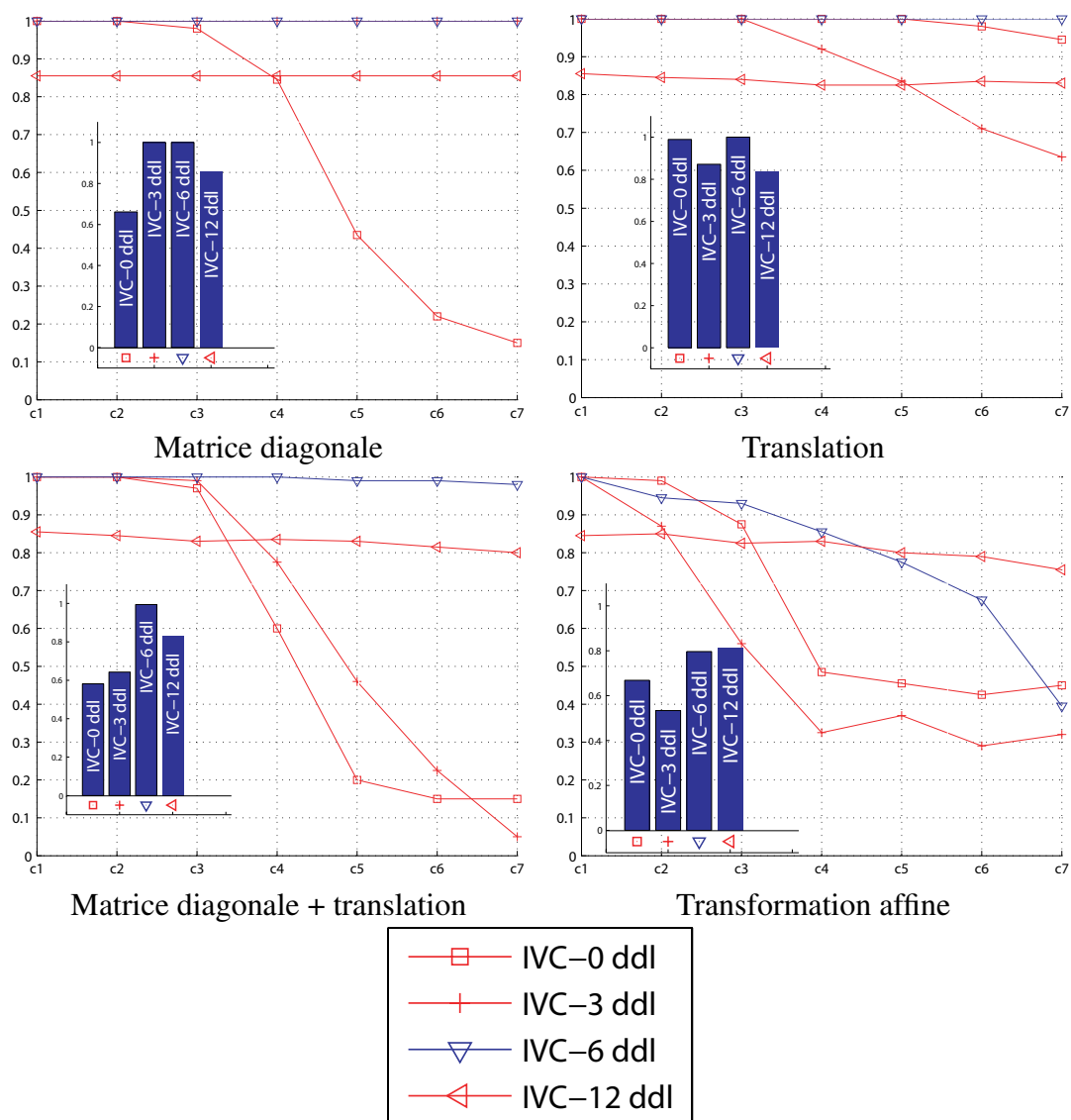


FIGURE 6.10 – La normalisation d’IVC sur la base ALOI (perturbations artificielles). Les conditions  $C_1$  à  $C_7$  sont décrites dans le tableau 6.1.

## CHAPITRE 6. RÉSULTATS EXPÉRIMENTAUX

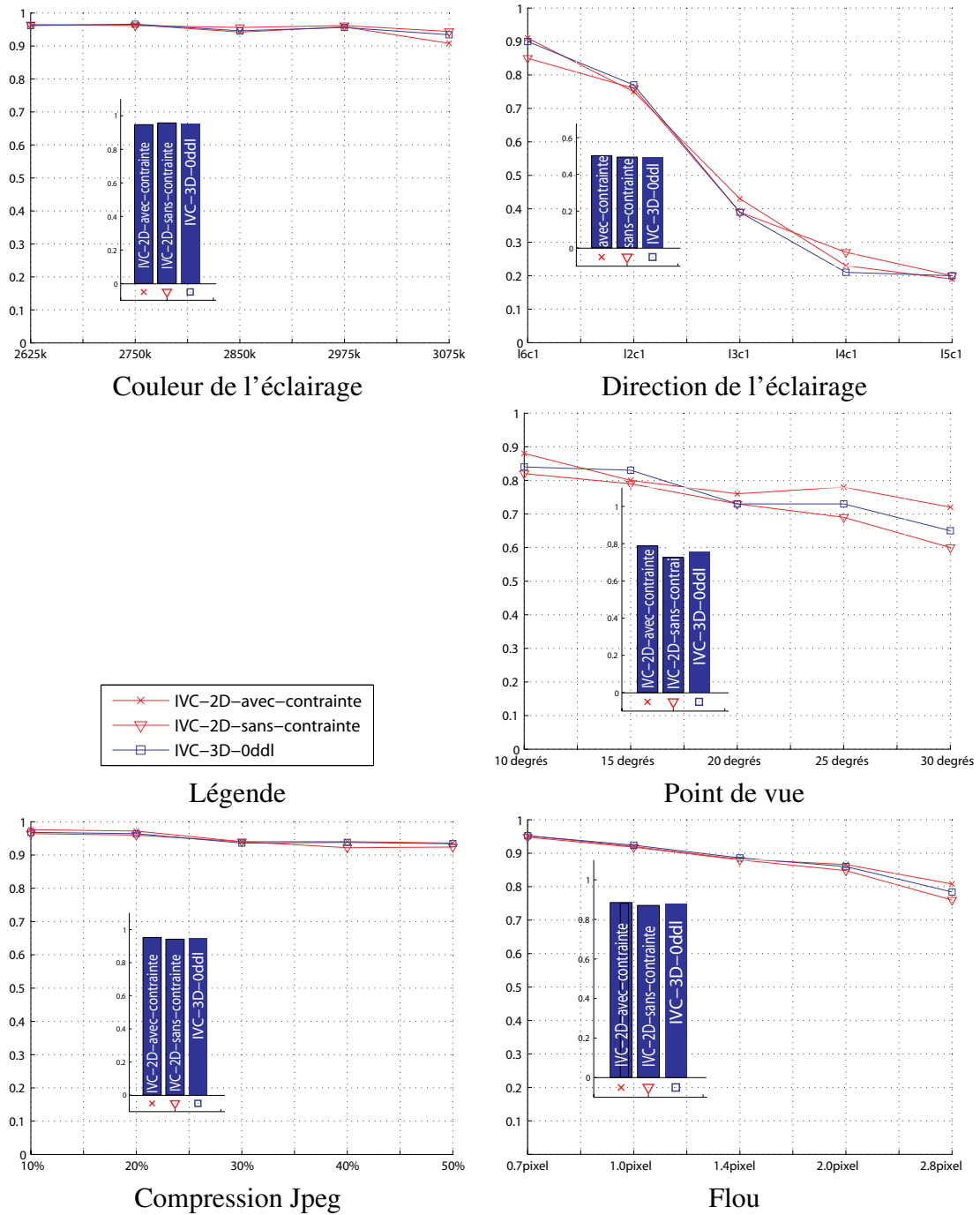


FIGURE 6.11 – IVC 3D et 2D avec et sans contrainte

## CHAPITRE 6. RÉSULTATS EXPÉRIMENTAUX

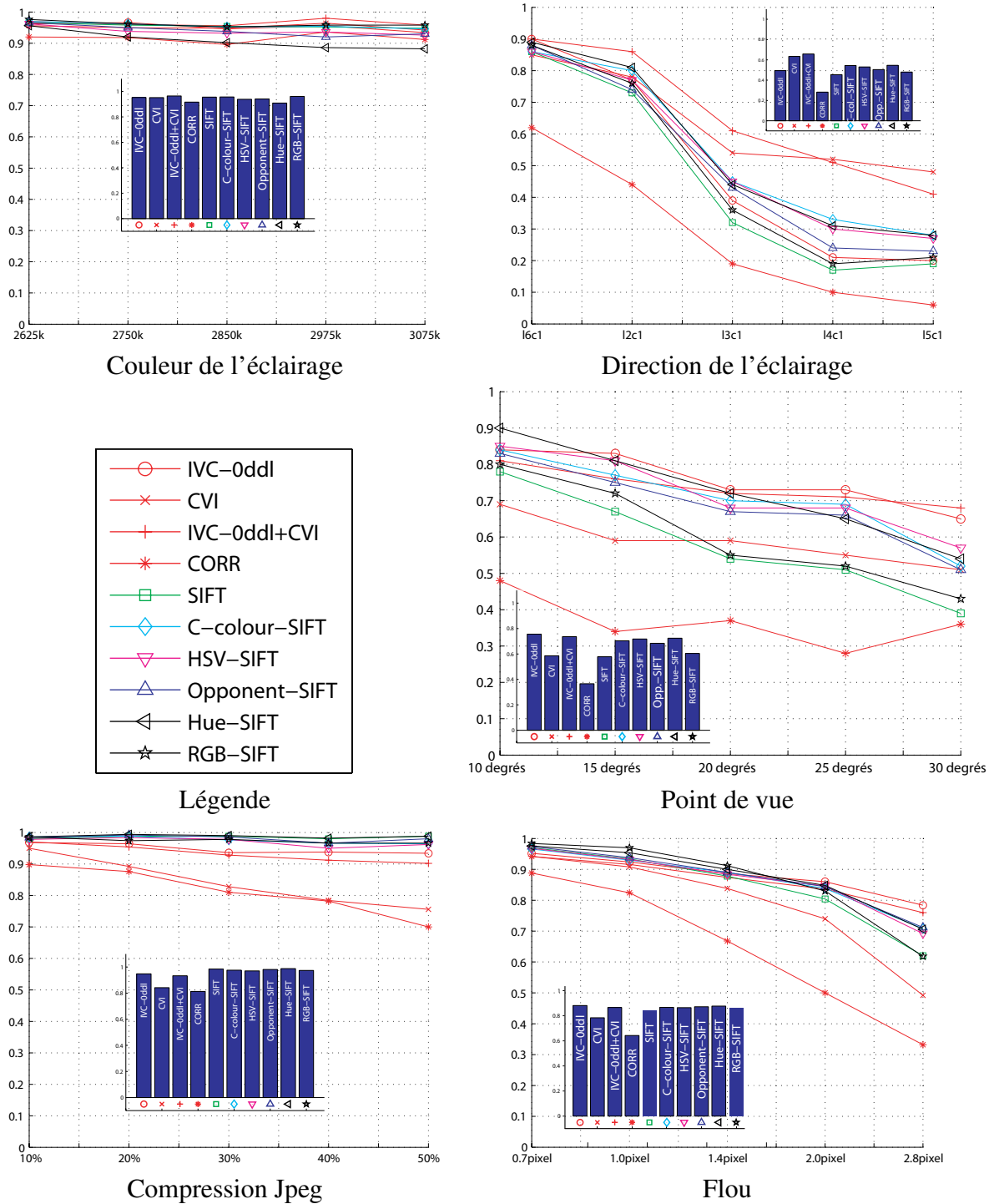


FIGURE 6.12 – Performances des descripteurs locaux sur la base ALOI



## CHAPITRE 6. RÉSULTATS EXPÉRIMENTAUX

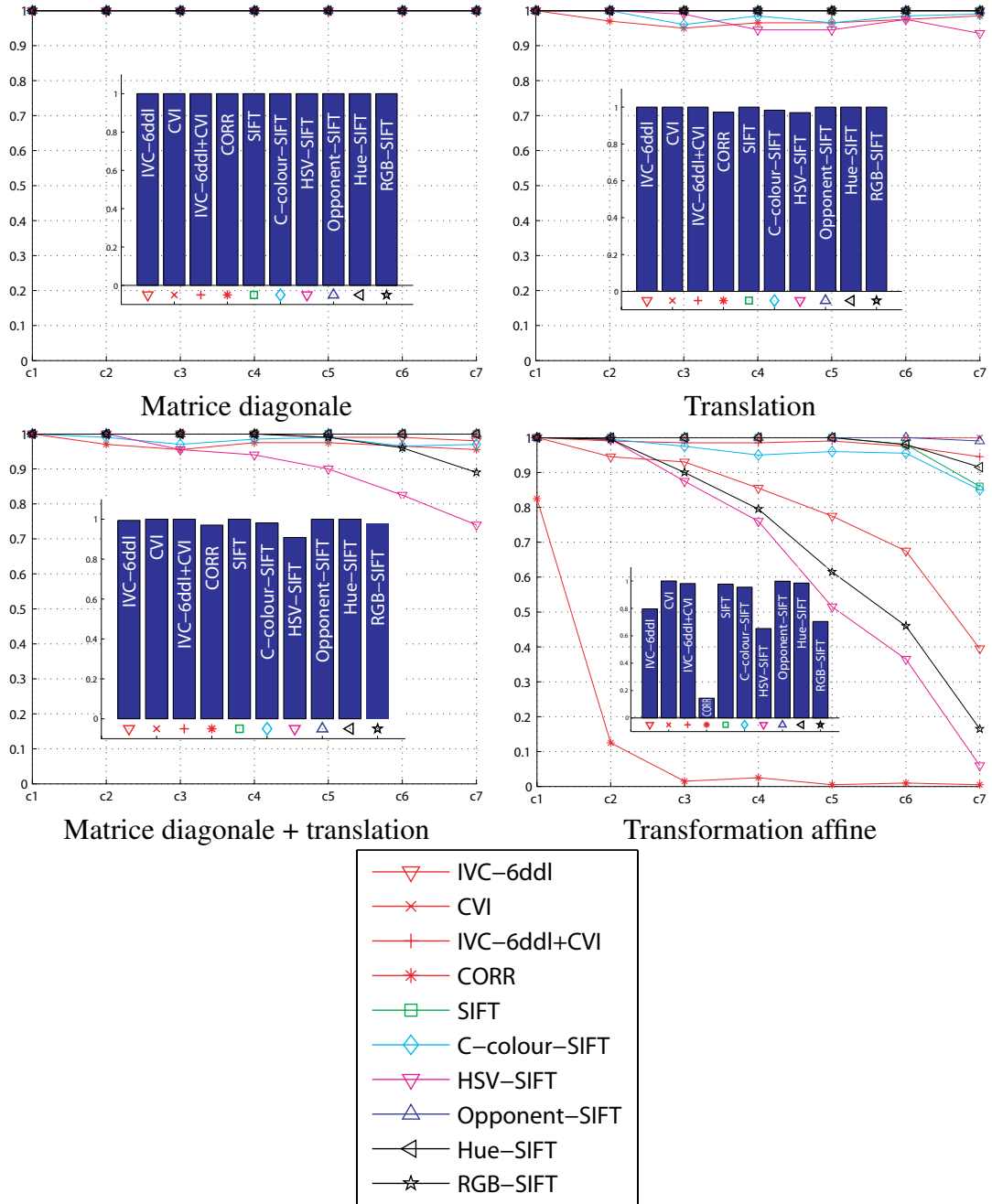


FIGURE 6.13 – Performances des descripteurs locaux sur la base ALOI (perturbations artificielles). Les conditions  $C_1$  à  $C_7$  sont décrites dans le tableau 6.1.

## 6.5 Classification d'objets

Dans le contexte de la classification d'objets, nous avons utilisé 4 bases différentes qui sont Flowers, Soccers, Birds et Pascal Voc 2010. Notre but ici n'est pas de fournir les meilleurs résultats de la littérature mais de comparer les performances de nos descripteurs avec ceux de la littérature dans des conditions identiques. Par conséquent, nous n'avons pas tenté d'affiner tous les paramètres qui ont un impact sur les résultats de classification. Ainsi, nous avons extrait une région de taille  $16 \times 16$  tous les 8 pixels pour les bases Flowers, Soccers et Birds et de taille  $12 \times 12$  tous les 6 pixels sur la base VOC 2010. Ceci est dû au fait que, pour la base VOC 2010, nous avons travaillé sur les fenêtres qui contiennent les objets et non sur les images globales. Les tailles étaient donc plus faibles. Pour chaque base, nous avons sélectionné aléatoirement 50.000 régions pour réduire le nombre de points dans l'espace des descripteurs. Ensuite, nous avons appliqué l'algorithme k-means pour extraire 200 mots visuels. Les images sont ensuite caractérisées par des histogrammes de mots visuels (sacs de mots). L'étape de classification est fondée sur les SVM avec un noyau d'intersection. Ceci nous permet d'évaluer la précision moyenne (MAP) sur chaque base comme illustré sur la Figure 6.14. Nous précisons que les échelles de ces figures sont différentes. Nous avons testé tous les descripteurs présentés précédemment ainsi que les descripteurs IVC utilisant différents nombres de degrés de liberté : de IVC-0ddl sans invariance photométrique à IVC-12ddl avec une invariance complète aux transformations affines appliquées dans l'espace couleur. Le descripteur CORR n'a pas été intégré à ces tests car quelques tests nous ont montré qu'il n'était pas adapté à la classification, certainement pour les raisons précisées ci-dessous.

La figure 6.14 nous montre que les descripteurs qui présentent des degrés d'invariance trop élevés (CVI, IVC-6ddl, IVC-12ddl) ne fournissent pas de très bons résultats en classification lorsque la couleur est une information importante. Ceci peut s'expliquer par le fait que l'invariance photométrique peut être atteinte par

l'utilisation de la méthode de sacs de mots. En effet, 200 mots visuels ont été sélectionnés pour chaque base et chaque descripteur extrait est associé au mot visuel le plus proche dans l'espace des descripteurs. Ainsi, si la couleur varie un peu d'une région à l'autre, l'association au mot visuel le plus proche risque de gommer cette variation. L'invariance devient alors inutile et nous savons que toute invariance inutile tend à diminuer le pouvoir discriminant. Cette faiblesse des descripteurs trop invariants est moins marquée pour la base VOC 2010 pour laquelle l'information de couleur n'apporte que peu d'informations par rapport à la forme.

Nous avons déjà précisé que :

- pour la base Soccers, la couleur est plus importante que la forme,
- pour la base Flowers, la couleur et la forme sont importantes,
- pour la base Birds, la couleur et la forme sont importantes, mais la forme est prédominante,
- pour la base VOC 2010, la couleur n'apporte pas beaucoup lorsqu'elle est ajoutée à la forme.

Nous remarquons que les spécificités de chaque base apparaissent dans les résultats. En effet, lorsque nous comparons les résultats des SIFT niveaux de gris et des SIFT couleur, nous pouvons en déduire si les informations de forme et de couleur sont importantes. Pour Soccers, les SIFT couleur ont des résultats de plus de 20% supérieurs aux SIFT niveaux de gris. Pour Flowers, l'augmentation est de 10%. Pour Birds et VOC 2010, l'augmentation moyenne est très faible.

Nous remarquons que le descripteur IVC-Oddl qui n'offre aucune invariance théorique fournit de très bons résultats de classification, et notamment sur les bases pour lesquelles la couleur est importante. Il permet en effet d'accroître de plus de 7% sur la base Flowers et de plus de 6% sur la base Soccers les performances du meilleur SIFT couleur (Hue-SIFT).

Le descripteur d'auto-similarités proposé dans le chapitre 5 n'a pas été considéré dans les tests ci-dessus. Nous proposons de le tester sur la base PASCAL VOC 2007 et de le comparer aux descripteurs de similarités locales [95] (BOLSS pour Bag Of Local Self-Similarities) et de similarités globales [24] (BOCS pour Bag Of Correlation Surfaces) introduits dans ce chapitre 5. Comme le soulignent Shechtman *et al.* [95] et Deselaers *et al.* [24], l'utilisation des auto-similarités pour la classification d'objets n'est pas suffisante en soi et fournit des résultats moins intéressants que les descripteurs locaux classiques, du fait de leur invariance très élevée aux variations photométriques. Jusqu'à aujourd'hui, elles sont davantage utilisées en complément des descripteurs locaux classiques. Notre objectif, encore une fois, n'est pas d'obtenir les meilleurs résultats de classification sur la base VOC 2007 de la littérature mais simplement de comparer dans les mêmes conditions, les approches d'auto-similarités de la littérature avec notre approche. Les résultats sont présentés dans la figure 6.15. Notre descripteur est appelé BO-RISS pour Bag Of Rotation Invariant Self Similarities.

Dans cette figure, nous pouvons constater que notre approche fournit de meilleurs résultats que les deux autres testées pour la plupart des classes de la base VOC 2007. De plus, la précision moyenne de notre approche est de 25% alors que celle des approches de la littérature est de 18%. Ces résultats très encourageants montrent bien que l'invariance à la rotation permet de retrouver des auto-similarités dans les images qui sont discriminantes dans le cadre de la classification d'objets.

## CHAPITRE 6. RÉSULTATS EXPÉRIMENTAUX

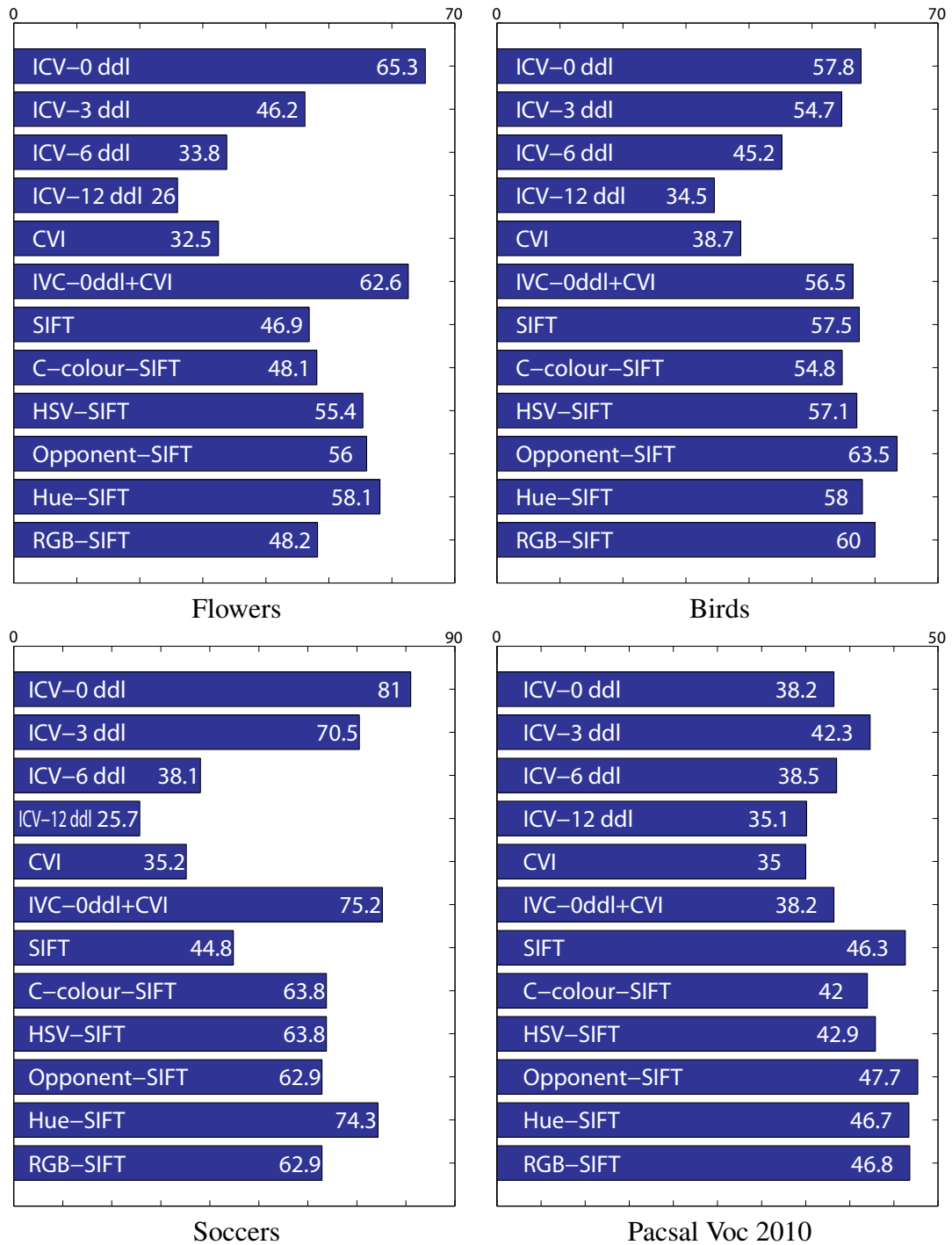


FIGURE 6.14 – Précision moyenne obtenue par les descripteurs pour la classification sur 4 bases

## CHAPITRE 6. RÉSULTATS EXPÉRIMENTAUX

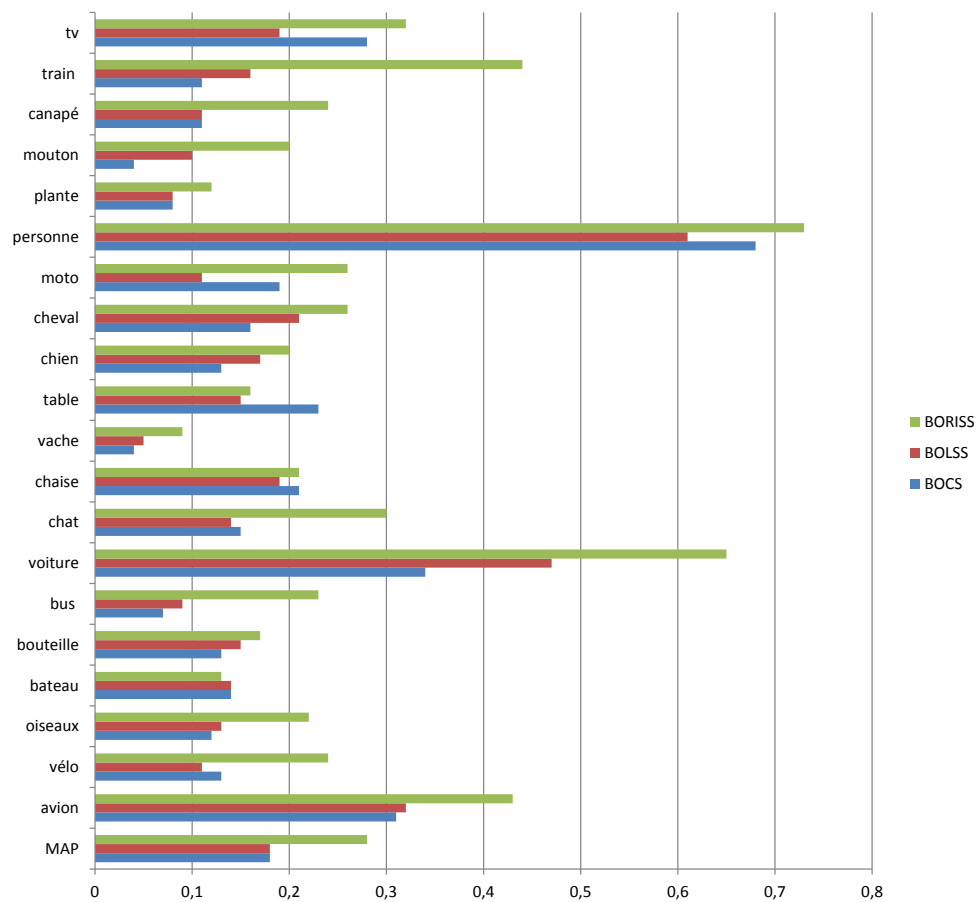


FIGURE 6.15 – Précisions obtenues par les descripteurs d’auto-similarités sur la base VOC 2007. MAP est le pourcentage de la précision moyenne.

## 6.6 Conclusion

Dans ce chapitre, nous avons exposé les performances obtenues par nos descripteurs et nous les avons comparées à celles des descripteurs locaux de la littérature. Tout d'abord, nous avons testé ces descripteurs dans le cadre de l'appariement de régions locales extraites d'images acquises sous des conditions très variables. Nous avons testé 9 conditions de variations différentes et nous avons pu constater que la théorie était validée par les résultats expérimentaux. Ainsi, nos descripteurs compacts fournissent des résultats supérieurs en moyenne à tous les autres descripteurs. Ceci s'explique par le fait que nos descripteurs ne sont pas fondés sur l'information de gradients locaux mais représentent les distributions spatiales des couleurs dans les régions. Cette information est très discriminante lorsqu'elle n'est pas sensible aux variations des conditions d'acquisition et nous avons montré que nos descripteurs parviennent à respecter le compromis entre pouvoir discriminant et degré d'invariance. De plus, nous avons montré que cette information peut être facilement résumée par quelques valeurs alors que les gradients locaux conduisent rarement à des descripteurs compacts.

Dans un second temps, nous avons testé nos descripteurs dans le contexte de la classification d'objets en utilisant une approche par sacs de mots. Une fois encore, nous avons montré que les descripteurs mis au point permettent d'obtenir de très bons résultats de classification sur des bases pour lesquelles la couleur est discriminante.





---

---

## CHAPITRE 7

---

# CONCLUSION GÉNÉRALE ET PERSPECTIVES

Dans le cadre de cette thèse nous nous sommes attachés aux propriétés d'invariance des descripteurs couleur locaux. Différents types de descripteurs invariants sont couramment employés dans des applications de reconnaissance, de catégorisation d'objets, d'indexation d'images, etc. Parmi les différentes propriétés d'invariance recherchées dans ce type d'applications, nous nous sommes plus particulièrement focalisés sur les propriétés d'invariance aux conditions d'observations.

De nombreux descripteurs locaux ont été introduits ces dix dernières années. Pour différencier ces différents descripteurs nous avons tout d'abord listé de manière structurée l'ensemble des hypothèses (propriétés d'invariance) sur lesquelles reposent ces descripteurs, puis nous avons établi une classification de ces descripteurs basée sur ces hypothèses. Nous avons également recensé les principales particularités qui caractérisent ces descripteurs, mettant en valeur leurs similitudes, leur degré de généralité et leur compacité. Nous avons focalisé cette étude sur les descripteurs couleur, ces derniers laissant présager une amélioration des perfor-

## CHAPITRE 7. CONCLUSION GÉNÉRALE ET PERSPECTIVES

---

mances par rapport aux descripteurs à niveaux de gris.

L'analyse des interactions qui existent entre les différents paramètres qui entrent en jeu dans la formation de la couleur nous a permis de proposer de nouveaux descripteurs locaux couleur. Ces descripteurs présentent des propriétés d'invariance aux conditions d'acquisition très diverses. En effet, dans le chapitre 3, nous avons proposé des descripteurs invariants aux transformations affines appliquées dans l'espace image comme dans l'espace couleur. Ces invariances permettent de couvrir une large partie des variations des conditions d'acquisition. Cependant, lorsque les variations photométriques conduisent à des transformations non linéaires des composantes couleur, le modèle affine ne fonctionne plus et nous avons proposé une approche qui exploite les mesures de rang dans le chapitre 4. Enfin, si les variations photométriques ne peuvent pas être modélisées par une fonction mathématique ou si le degré de généralité requis est très important (en classification par exemple), nous avons plutôt proposé d'exploiter les distributions spatiales des auto-similarités dans les images.

Ainsi, dans le chapitre 3, nous avons proposé d'appliquer des transformations affines qui permettent de transformer les coordonnées spatiales des pixels en coordonnées couleur et inversement. Cette idée originale d'application de transformation affine entre deux espaces différents nous a permis de mettre au point un ensemble de descripteurs locaux dont les propriétés d'invariance sont diverses. D'une part, la transformation de l'espace image à l'espace couleur nous fournit un descripteur très peu sensible aux variations géométriques dans l'espace image, auquel nous avons ajouté une étape de normalisation photométrique afin de contrôler son invariance aux transformations qui peuvent perturber les couleurs. Nous avons montré qu'en contraignant plus ou moins cette étape de normalisation en jouant sur le nombre de degrés de liberté de la fonction associée, nous pouvons contrôler le degré d'invariance du descripteur et ainsi s'adapter à l'application considérée. D'autre part, la transformation de l'espace couleur vers l'espace image conduit à un descripteur invariant à toute transformation affine appliquée dans l'espace

## CHAPITRE 7. CONCLUSION GÉNÉRALE ET PERSPECTIVES

---

couleur. De plus comme nous utilisons une décomposition en sous-fenêtre, ce descripteur peu sensible aux variations spatiales de l'éclairage. Les descripteurs présentés dans ce chapitre représentent à la fois les couleurs de la région considérée et leur répartition spatiale. Les tests menés dans le chapitre 6 ont mis en évidence le pouvoir discriminant de ces descripteurs compacts. En effet, ils fournissent de très bons résultats dans le cadre de l'appariement de régions comme pour la classification d'images.

Dans le chapitre 4, nous avons abordé les problèmes de variations photométriques qui sont modélisés par une variation non-linéaire des composantes couleur. Dans ce cas, l'invariance aux transformations affines n'est plus suffisante et nous avons proposé d'exploiter les mesures de rang des pixels. Dans le cadre d'une description locale, nous tenions à insérer une information quant à la répartition spatiale des mesures de rang des pixels dans la région considérée. Pour cela, nous avons proposé d'exploiter la mesure de corrélations de rang de Kendall pour extraire les corrélations entre sous-ensembles disjoints de pixels. Dans ce but, il a été nécessaire d'étendre la définition de ces corrélations à des sous-ensembles disjoints. De plus, nous avons proposé une approche qui permet d'extraire très rapidement ces corrélations à partir d'histogrammes couleur. Le descripteur mis au point dans ce chapitre représente donc les corrélations entre composantes couleur des pixels appartenant à des sous-régions différentes. Il caractérise donc les répartitions spatiales des mesures de rang tout en prenant en compte plusieurs composantes couleur.

La problématique abordée dans le chapitre 5 est un peu différente puisqu'il s'agit de caractériser le contenu d'une image, non par un ensemble de caractéristiques locales mais par la distribution spatiale des similarités entre régions. Cette approche permet de reconnaître des objets qui ne partagent pas forcément des caractéristiques locales communes. Différentes approches ont été proposées dans ce contexte mais nous avons mis en évidence qu'aucune d'elle n'est invariante à la rotation et cette sensibilité limite le nombre d'auto-similarités détectées. Nous

avons donc proposé une méthode pour extraire des auto-similarités entre régions quelles que soient les orientations de ces régions dans l'espace image. Une fois que ces auto-similarités sont détectées, nous avons mis au point une structure *3D* qui permet de résumer leur distribution spatiale locale relative. Nous avons montré dans le chapitre 6 que cette structure est très pertinente dans le cadre de la classification d'images puisqu'elle permet d'améliorer de manière conséquente les résultats des approches classiques.

Les travaux réalisés au cours de cette thèse nous ouvrent un large champs de perspectives.

D'une part, notre étude sur les invariants couleur a révélé qu'une large part des invariances sont obtenues par dérivées spatiales ou rapports entre composantes de pixels voisins. Les valeurs déduites de ces analyses très locales sont donc sensibles aux variations des composantes des pixels. Cependant, la plupart des images sont acquises par des capteurs mono-CCD pour lesquels un seul niveau de composante est connu pour chaque pixel, les 2 autres niveaux étant déduits par des algorithmes de démosaïçage. Une étude intéressante pourrait être d'analyser l'impact des algorithmes de démosaïçage sur les invariants colorimétriques obtenus.

Ensuite, nous avons proposé un ensemble de descripteurs locaux dont les propriétés d'invariance sont très diverses avec l'idée que le choix du descripteur doit être fait en fonction de l'application. Le descripteur IVC, à lui seul, peut être adapté à plusieurs conditions en jouant sur le nombre de degrés de liberté de sa fonction de normalisation. Cette fonction est celle qui permet de modifier les positions des coins des sous-fenêtres qui constituent la région à caractériser par rapport aux positions des coins de la région complète. Cette normalisation est réalisée avant de transmettre le descripteur mais nous soutenons l'idée qu'il serait possible d'ajouter les positions des coins de la région complète comme un en-tête du descripteur afin d'adapter le nombre de degrés de liberté aux couples de régions comparées. Ceci permettrait de ne pas avoir à faire le choix de la normalisation avant de commencer la comparaison entre régions.

Dans ce manuscrit nous n'avons pas abordé la question relative à l'extraction des régions d'intérêt, partant du fait que ces régions ont été extraites et qu'il s'agit de les caractériser. Les régions extraites dans cette thèse l'ont été avec le détecteur Harris-affine [72] qui est invariant aux transformations affines appliquées dans l'espace image. Cependant ce descripteur n'exploite pas l'information de couleur pour extraire ces régions et nous proposons d'exploiter les caractéristiques de notre descripteur CVI pour extraire des régions couleur dans les images. En effet, ce descripteur est complètement invariant aux transformations affines appliquées dans l'espace couleur et consiste à transformer les distributions couleur sur les positions 2D dans l'image. Si nous considérons des sphères dans l'espace couleur et que nous appliquons cette transformation, nous voyons apparaître des ellipses qui couvrent une partie des pixels de la région et dont les formes, orientations, échelles et positions sont invariantes aux transformations géométriques et photométriques. Nous envisageons d'exploiter ces propriétés pour extraire des régions d'intérêt dans les images.

La description par auto-similarités est une approche très récente. L'un de ses points faibles est l'invariance à l'échelle car les tailles des régions et des voisinages étant fixées, les objets décrits doivent avoir des tailles similaires dans les images. Notre approche s'adapterait particulièrement bien à une analyse multi-échelle puisque nous sommes capables d'évaluer des similarités entre régions de tailles différentes. En effet, il suffirait d'éliminer les termes de changement d'échelle dans les matrices de transformation  $T_{IVC}$  tout comme nous l'avons fait avec les termes de rotation pour obtenir l'invariance à la rotation.

Le calcul d'auto-similarités multi-échelle est un point clé de la compression fractale qui code les images par une liste de transformation affine qui relie des régions de tailles différentes dans les images. La décompression fractale consiste à appliquer itérativement ces transformations affines à une image quelconque. Une perspective de notre travail pourrait être d'adapter notre approche à l'extraction d'auto-similarités multi-échelle pour la compression fractale. A notre connais-

## CHAPITRE 7. CONCLUSION GÉNÉRALE ET PERSPECTIVES

---

sance, les auto-similarités utilisées pour la compression fractale ne sont pas invariantes à la rotation, limitant ainsi la qualité de l'image reconstruite. Notre approche trouverait tout son intérêt dans cette problématique.

---

# BIBLIOGRAPHIE

- [1] Base caltech. URL <http://www.vision.caltech.edu/html-files/archive.html>.
- [2] Base graz02. URL [http://www.emt.tugraz.at/~pinz/data/GRAZ\\_02/](http://www.emt.tugraz.at/~pinz/data/GRAZ_02/).
- [3] Birds dataset. URL [http://www-cvr.ai.uiuc.edu/ponce\\$\\_\\$grp/data/](http://www-cvr.ai.uiuc.edu/ponce$_$grp/data/).
- [4] Color sift page from j.m. geusebroek. URL <http://staff.science.uva.nl/~mark/downloads.html>.
- [5] Flower dataset. URL <http://www.robots.ox.ac.uk/vgg/research/flowers/>.
- [6] Pascal voc challenge 2010. URL <http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2010/index.html>.
- [7] Soccer dataset. URL <http://lear.inrialpes.fr/data>.
- [8] Visual geometry group, katholieke universiteit leuven. URL

<http://www.robots.ox.ac.uk/~vgg/research/affine/descriptors.html>.

- [9] Web page of j. van de weijer. URL <http://cat.cvc.uab.es/~joost/software.html>.
- [10] A.E. Abdel-Hakim et A.A. Farag. Csift : A sift descriptor with color invariant characteristics. volume 2, pages 1978–1983, 2006.
- [11] C. Ancuti et P. Bekaert. Sift-cch : Increasing the sift distinctness by color co-occurrence histograms. Dans *Procs. of the 5th International Symposium on Image and Signal Processing and Analysis*, pages 130–135, Istambul, Turkey, 2007.
- [12] S. Battiato, G. Gallo, G. Puglisi et S. Scellato. Sift features tracking for video stabilization. pages 825–830, Los Alamitos, CA, USA, 2007. IEEE Computer Society.
- [13] H. Bay, A. Ess, T. Tuytelaars et L. Van Gool. Surf : Speeded up robust features. *Computer Vision and Image Understanding*, 110 (3):346–359, 2008.
- [14] P. Beckmann et A. Spizzichino. *The scattering of electromagnetic waves from rough surfaces. A Survey of methods for colour image indexing and retrieval in image databases*. Artech House Inc, 2<sup>nd</sup> edition, 1987.
- [15] A. Bosch, A. Zisserman et X. Munoz. Scene classification via pls. pages 517–530, 2006.
- [16] G. Buchsbaum. A spatial processor model for object colour perception. *Jour. of the Franklin Institute*, 310:1–26, 1980.
- [17] G.J. Burghouts et J.-M. Geusebroek. Performance evaluation of local colour invariants. *Computer Vision and Image Understanding*, 113 (1):48–62, 2009.



- [18] P. Chang et J. Krumm. Object recognition with color cooccurrence histograms. Dans *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, page 504, 1999.
- [19] K. Chatfield, J. Philbin et A. Zisserman. Efficient retrieval of deformable shape classes using local self-similarities. Dans *NORDIA workshop in conjunction with ICCV 2009*, 2009.
- [20] Xin Chen, Xiaohua Hu et Xiaojiong Shen. Spatial weighting for bag-of-visual-words and its application in content-based image retrieval. Dans *Advances in Knowledge Discovery and Data Mining*, volume 5476 de *Lecture Notes in Computer Science*, pages 867–874. 2009.
- [21] D. M. Chu et A. W. M. Smeulders. Color invariant surf in discriminative object tracking. Dans *ECCV Workshop on Color and Reflectance in Imaging and Computer Vision*, 2010.
- [22] G. Ciocca, D. Marini, A. Rizzi, R. Schettini et S. Zuffi. On pre-filtering with retinex in color image retrieval. Dans *Procs. of the SPIE Conf. on Internet Imaging II*, volume 4311, pages 140–147, 2001.
- [23] A.B. Dahl et H. Aanaes. Effective image database search via dimensionality reduction. pages 1–6, 2008.
- [24] T. Deselaers et V. Ferrari. Global and efficient self-similarity for object classification and detection. Dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, 2010.
- [25] M.S. Drew, J. Wei et Z.N. Li. On illumination invariance in color object recognition. *Pattern Recognition*, 31(8):1077–1087, 1998.
- [26] I. Elsayad, J. Martinet, T. Urruty et C. Djeraba. A new spatial weighting scheme for bag-of-visual-words. Dans *Proceedings of the International*

- Workshop on Content-Based Multimedia Indexing (CBMI 2010)*, pages 1–6, 2010.
- [27] A.A. Farag et A.E. Abdel-Hakim. Detection, categorization and recognition of road signs for autonomous navigation. Dans *Procs. of Advanced Concepts in Intelligent Vision Systems*, pages 125–130, Brussel, Belgium, 2004.
- [28] G. Finlayson, S. Hordley, G. Schaefer et G. Y. Tian. Illuminant and device invariant colour using histogram equalisation. *Pattern Recognition*, 38:179–190, 2005.
- [29] G. D. Finlayson, M. S. Drew et B. V. Funt. Spectral sharpening : sensor transformations for improved color constancy. *Jour. of the Optical Society of America*, 11(A):1553–1563, 1994.
- [30] G. D. Finlayson, S. D. Hordley et R. Xu. Convex programming colour constancy with a diagonal-offset model. Dans *Proc. IEEE International Conference on Image Processing*, pages 948–951, 2005.
- [31] G.D. Finlayson, S.S Chatterjee et B.V. Funt. Color angle invariants for object recognition. Dans *Procs. of the 3<sup>rd</sup> IS&T/SID Color Imaging Conf.*, pages 44–47, 1995.
- [32] G.D. Finlayson, M.S. Drew et B.V. Funt. Color constancy : enhancing von kries adaption via sensor transformations. Dans J. P. Allebach & B. E. Rogowitz, éditeur, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 1913, pages 473–484, september 1993.
- [33] G.D. Finlayson, M.S. Drew et B.V. Funt. Color constancy : Generalized diagonal transforms suffice. *Jour. of the Optical Society of America*, 11 (A):3011–3020, 1994.

- [34] G.D. Finlayson, M.S. Drew et C. Lu. Intrinsic images by entropy minimization. Dans *Procs.of the European Conf. on Computer Vision*, pages 582–595, 2004.
- [35] G.D. Finlayson et S.D. Hordley. Colour constancy at a pixel. *Jour. of the Optical Society of America*, 18(2):253–264, 2001.
- [36] G.D. Finlayson, S.D. Hordley et P.M. Hubel. Color by correlation : a simple, unifying framework for color constancy. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(11):1209–1221, 2001.
- [37] G.D. Finlayson, B. Schiele et J.L. Crowley. Comprehensive colour image normalization. *Lecture Notes in Computer Science*, 1406:475–490, 1998.  
URL [citeseer.nj.nec.com/finlayson98comprehensive.html](http://citeseer.nj.nec.com/finlayson98comprehensive.html).
- [38] Graham D. Finlayson et Elisabetta Trezzi. Shades of gray and colour constancy. Dans *Proc. Color Imaging Conference*, pages 37–41, 2004.
- [39] B. Funt, V. C. Cardei et K. Barnard. Method of estimating chromaticity of illumination using neural networks. Dans *United States Patent*, volume 5,907,629, USA, 1999.
- [40] B.V. Funt et G.D. Finlayson. Color constant color indexing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(5):522–529, 1991.
- [41] B.V. Funt et G.D. Finlayson. Color constant color indexing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(5):522–529, 1995.
- [42] R. Gershon, A. D. Jepson et J. K. Tsotsos. From [r,g,b] to surface reflectance : computing color constant descriptors in images. *Perception*, pages 755–758, 1988.

- [43] J. M. Geusebroek, G. J. Burghouts et A. W. M. Smeulders. The Amsterdam library of object images. *Int. J. Comput. Vision*, 61(1):103–112, 2005.
- [44] J.-M. Geusebroek, R. van den Boomgaard, A. W. M. Smeulders et A. Dev. Color and scale : The spatial structure of color images. Dans *Procs.of the European Conf. on Computer Vision*, pages 331–341, 2000.
- [45] J. M. Geusebroek, R. van den Boomgaard, A. W. M. Smeulders et H. Geerts. Color invariance. *IEEE Trans. Pattern Anal. Machine Intell.*, 23(12):1338–1350, 2001.
- [46] J.M. Geusebroek. Compact object descriptors from local colour invariant histograms. Dans *British Machine Vision Conference*, volume 3, pages 1029–1038, 2006.
- [47] T. Gevers et A.W.M. Smeulders. Color-based object recognition. *Pattern Recognition*, 32:453–464, 1999.
- [48] T. Gevers et A.W.M. Smeulders. Content-based image retrieval by viewpoint-invariant color indexing. *Image and Vision Computing*, 17:475–488, 1999.
- [49] G.Healey et D. Slater. Global color contancy :recognition of objects by use of illumination invariant properties of color distributions. *Jour. of the Optical Society America*, 11(11):3003–3010, 1995.
- [50] T. Goedemé, T. Tuytelaars et L. Van Gool. Omnidirectional sparse visual path following with occlusion-robust feature tracking. Dans *in : 6th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras, OMNIVIS05, in Conjunction with ICCV 2005*, 2005.
- [51] P. Gros. Color illumination models for image matching and indexing. Dans *Procs. of the 15<sup>th</sup> Int. Conf. on Pattern Recognition*, volume 3, pages 576–579, Barcelona, Spain, 2000.

- [52] D. Hegazy et J. Denzler. Boosting colored local features for generic object recognition. *Pattern Recognition and Image Analysis*, 18 (2):323–327, 2008.
- [53] L. Hu, S.Q.A. Jiang, Q.M. Huang et W. Gao. People re-detection using adaboost with sift and color correlogram. pages 1348–1351, 2008.
- [54] J. Huang, S. Ravi Kumar, M. Mitra, W. Zhu et R. Zabih. Image indexing using color correlogram. Dans *Procs. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 762–768, 1997.
- [55] International Radio Consultative Committee (ITU). Basic parameter values for the hdtv standard for the studio and for international programme exchange. Rapport technique 709-2, CCIR Recommendation, 1990.
- [56] D.B. JUDD, D.L. MACADAM, G. WYSZECKI, H.W. BUDDE, H.R. CONDIT, S.T. HENDERSON et J.L. SIMONDS. Spectral distribution of typical daylight as a function of correlated color temperature. *Journal of Optical Society of America*, 54(8):1031–1036, Aug 1964.
- [57] Y. Ke et R. Sukthankar. Pca-sift : A more distinctive representation for local image descriptors. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2:506–513, 2004. ISSN 1063-6919.
- [58] M.G. Kendall. *Rank Correlation Methods*. Griffin, 1970.
- [59] Y. Kenji, Y. Tomohiro, F. Toshiaki, T. Masayuki et S. David. Colour correction for multiple-camera system by using correspondences. *Journal of the Institute of Image Information and Television Engineers*, 61(2):213 – 222, 2007.
- [60] F.S. Khan, J. van de Weijer et M. Vanrell. Top-down color attention for object recognition. Dans *Procs. of the International Conference on Computer Vision*, pages 979–986, Japan, 2009.

- [61] G. Klinker, S. Shafer et T. Kanade. A physical approach to color image understanding. *Inter. journ. of Computer Vision*, 4(1):7–38, 1991.
- [62] D. Koubaroulis, J. Matas et J. Kittler. Illumination invariant object recognition using the MNS method. Dans *Procs. of the 10<sup>th</sup> European Signal Processing Conf.*, pages 2173–2176, Tampere University of Technology, Finland, september 2000.
- [63] P. Kubelka. New contribution to the optics of intensity light-scattering materials, part i. *J. Optical Soc. Am. A*, 38(5):448–457, 1948.
- [64] E. Land. The retinex theory of color constancy. *Scientific American*, pages 108–129, 1977.
- [65] R. Lenz, L.V. Tran et P. Meer. Moment based normalization of color images. Dans *Procs. of the IEEE Workshop on Multimedia Signal Processing*, pages 129–132, 1999.
- [66] Thomas Leung et Jitendra Malik. Representing and Recognizing the Visual Appearance of Materials using Three-dimensional Textons. *Int. J. Comput. Vision*, 43(1):29–44, juin 2001. ISSN 0920-5691. URL <http://dx.doi.org/10.1023/A:1011126920638>.
- [67] J. Li et N. M. Allinson. A comprehensive review of current local features for computer vision. *Neurocomput.*, 71(10-12):1771–1787, 2008. ISSN 0925-2312.
- [68] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [69] R. H. Luke, J. M. Keller et J. Chamorro-Martinez. Extending the scale invariant feature transform descriptor into the color domain. *Proceedings of the ICGST International Journal on Graphics, Vision and Image Processing, GVIP*, 08:35–43, 2008.

- [70] L.T. Maloney. Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *Journal of Optical Society of America*, 3(10):1673–1683, Oct 1986.
- [71] J. Matas, D. Koubaroulis et J. Kittler. Colour image retrieval and object recognition using the multimodal neighbourhood signature. Dans *Procs.of the European Conf. on Computer Vision*, pages 48–64, Berlin, Germany, 2000.
- [72] K. Mikolajczyk et C. Schmid. Scale & affine invariant interest point detectors. *Int. J. Comput. Vision*, 60:63–86, October 2004.
- [73] K. Mikolajczyk et C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Machine Intell.*, 27:1615–1630, 2005.
- [74] Krystian Mikolajczyk, Tinne Tuytelaars, Cordelia Schmid, Andrew Zisserman, J. Matas, F. Schaffalitzky, T. Kadir et L. Van Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1/2):43–72, 2005. URL <http://lear.inrialpes.fr/pubs/2005/MTSZMSKG05>.
- [75] F. Mindru, L.J. Van Gool et T. Moons. Model estimation for photometric changes of outdoor planar color surfaces caused by changes in illumination and viewpoint. Dans *Proceedings of the 16 th International Conference on Pattern Recognition*, volume 1 de *ICPR '02*, pages 620–623, Washington, DC, USA, 2002. IEEE Computer Society.
- [76] F. Mindru, T. Moons et L. van Gool. Recognizing color patterns irrespective of viewpoints and illuminations. Dans *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 368–373, 1999.
- [77] F. Mindru, T. Tuytelaars, L. Van Gool et T. Moons. Moment invariants for

- recognition under changing viewpoint and illumination. *Computer Vision and Image Understanding*, 1(3):3–27, 2004.
- [78] F. Mindru, T. Tuytelaars, L. Van Gool et T. Moons. Moment invariants for recognition under changing viewpoint and illumination. *Comput. Vis. Image Underst.*, 94(1-3):3–27, 2004.
- [79] J.M. Morel et G. Yu. Asift : A new framework for fully affine invariant image comparison. *SIAM Journal of Imaging Science*, 2:438–469, April 2009. ISSN 1936-4954.
- [80] D. Muselet. *Reconnaissance automatique d'objets sous éclairage non contrôlé par analyse d'images couleur*. Thèse de doctorat, Université Scientifique et Technologique de Lille, juillet 2005.
- [81] D. Muselet et A. Trémeau. Rank correlation as illumination invariant descriptor for color object recognition. Dans *Procs. of the Int. Conf. on Image Processing*, San Diego (California), 2008.
- [82] D. Muselet et L. Macaire. Fuzzy spatial ranks for object recognition across illumination changes. *Multimedia and Expo, IEEE International Conference on*, 0:985–988, 2006.
- [83] D. Muselet et L. Macaire. Combining color and spatial information for object recognition across illumination changes. *Pattern Recognition Letters*, 28 (10):1176–1185, 2007.
- [84] D. Muselet et L. Macaire. Combining color and spatial information for object recognition across illumination changes. *Pattern Recognition Letters*, 28 (10):1176–1185, 2007.
- [85] Sarif Kumar Naik et C. A. Murthy. Distinct multicolored region descriptors for object recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29:1291–1296, July 2007.



- [86] S.K. Nayar et R.M. Bolle. Reflectance based object recognition. *Inter. journ. of Computer Vision*, 17(3):219–240, 1996.
- [87] M.-E. Nilsback et A. Zisserman. Automated flower classification over a large number of classes. Dans *Proceedings of the Indian Conference on Computer Vision, Graphics Image Processing (ICVGIP 2008)*, pages 722–729, 2008.
- [88] J. P. S. Parkkinen, J. Hallikainen et T. Jaaskelainen. Characteristic spectra of munsell colors. *Journal of Optical Society of America*, 6(2):318–322, Feb 1989.
- [89] G. Qiu. Indexing chromatic and achromatic patterns for content-based colour image retrieval. *Pattern Recognition*, 35(8):1675–1686, 2002.
- [90] P. Quelhas et J.M. Odobez. Natural scene image modeling using color and texture visterms. Dans *Procs. of Conference on Image and Video Retrieval*, pages 411–421, Phoenix, USA, 2006.
- [91] P. Schügerl, R. Sorschag, W. Bailer et G. Thallinger. Object re-detection using sift and mpeg-7 color descriptors. Dans *Procs. of the International Workshop on Multimedia Content Analysis and Mining*, pages 305–314, 2007.
- [92] S. Se et P. Jasiobedzki. Photo-realistic 3d model reconstruction. Dans *Proceedings of IEEE International Conference on Robotics and Automation*, pages 3076–3082, 2006.
- [93] M. Seifi, X. Song, D. Muselet et A. Trémeau. Color texture classification across illumination changes. Dans *European Conference on Colour in Graphics, Imaging and Vision*, pages 332–337, Joensuu, Finland, 2010.
- [94] S. A. Shafer. Using color to separate reflection components. *Color Research and Application*, 10(4):210–218, 1985.

- [95] E. Shechtman et M. Irani. Matching local self-similarities across images and videos. Dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, 2007.
- [96] Y. Tsin, R.T. Collins, V. Ramesh et T. Kanade. Bayesian color constancy for outdoor object recognition. Dans *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 1132–1139, 2001.
- [97] Tinne Tuytelaars et Krystian Mikolajczyk. Local invariant feature detectors : A survey. *Foundations and Trends in Computer Graphics and Vision*, 3(3):177–280, 2008.
- [98] K.E.A. van de Sande, T. Gevers et C.G.M. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32:1582–1596, 2010.
- [99] J. van de Weijer, T. Gevers et J.-M. Geusebroek. Edge and corner detection by photometric quasi-invariants. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(4):625–630, 2005.
- [100] J. van de Weijer, T. Gevers et A.W.M. Smeulders. Robust photometric invariant features from the colour tensor. *IEEE Trans. on Image Processing*, 15(1):118–127, 2006.
- [101] J. van de Weijer et C. Schmid. Coloring local feature extraction. Dans *Procs. of the ninth European Conference on Computer Vision*, volume 3954, pages 334–348, Graz, Austria, 2006.
- [102] J. van de Weijer et C. Schmid. Applying color names to image description. Dans *Procs. of the IEEE Int. Conf. on Image Processing*, volume 3, pages 493–496, San Antonio (USA), 2007.

- [103] D. A. Rojas Vigo, F. S. Khan, J. van de Weijer et T. Gevers. The impact of color on bag-of-words based object recognition. Dans *International Conference on Pattern Recognition*, pages 1549–1553, 2010.
- [104] J. Vogel et B. Schiele. A semantic typicality measure for natural scene categorization. Dans Carl Edward Rasmussen, Heinrich H. Bülthoff, Bernhard Schölkopf et Martin A. Giese, éditeurs, *Pattern Recognition*, volume 3175 de *Lecture Notes in Computer Science*, pages 195–203. Springer Berlin / Heidelberg, 2004.
- [105] G. West et M.H. Brill. Necessary and sufficient conditions for von kries chromatic adaptation to give color constancy. *Journal of Mathematical Biology*, 15:249–258, 1982. ISSN 0303-6812. URL <http://dx.doi.org/10.1007/BF00275077>.
- [106] P. Wu, L. Kong, X. Li et K. Fu. A hybrid algorithm combined color feature and keypoints for object detection. Dans *Procs. of the 3rd IEEE Conference on Industrial Electronics and Applications*, pages 1408–1412, Singapore, 2008.
- [107] P. Wu, L. Kong, F. Zhao et X. Li. Particle filter tracking based on color and sift features. Dans *Procs. of the International Conference on Audio, Language and Image Processing*, Shanghai, 2008.
- [108] Z. Yi, C. Zhiguo et X. Yang. Multi-spectral remote image registration based on sift. *Electronics Letters*, 44(2):107–108, 2008.
- [109] D. Zhang, W. Wang, W. Gao et S. Jiang. An effective local invariant descriptor combining luminance and color information. Dans *Procs. of IEEE International Conference on Multimedia and Expo*, pages 1507–1510, Beijing (China), 2007.

- [110] H. Zhou, Y. Yuan et C. Shi. Object tracking using sift features and mean shift. *Computer Vision and Image Understanding*, 113(3):345 – 352, 2009. Special Issue on Video Analysis.

---

## TABLE DES FIGURES

1.1	Conditions d'acquisition non contrôlées dans le cadre de la recherche d'objets. . . . .	14
1.2	yy-ss . . . . .	19
2.1	Réflexion diffuse et spéculaire d'un rayon lumineux. Chaque élément est assujetti à diverse hypothèses détaillées dans ce chapitre.	24
3.1	Une ellipse détectée autour d'un pixel dans une image et le rectangle $N \times M$ qui la contient. . . . .	86
3.2	L'espace image $2D$ et l'espace couleur $3D$ . L'idée consiste à déterminer la 'meilleure' transformation affine de l'espace image vers l'espace couleur (IVC) ou de l'espace couleur vers l'espace image (CVI). Pour IVC (CVI, resp.) l'espace origine $\mathcal{O}$ est l'espace image (couleur, resp.) et l'espace destination $\mathcal{D}$ est l'espace couleur (image, resp.). . . . .	87
3.3	Transformation de l'espace image vers l'espace couleur du rectangle détecté dans la figure 3.2. Se référer au texte pour les explications. . . . .	91

3.4	Trois exemples de propriété d'invariance d'une transformation affine $T_{OVD}$ . A chaque point correspond un seul et unique symbole dans les espaces origine et destination. A partir de ces paires, nous évaluons la transformation affine qui permet de passer des positions d'origine (première colonne) aux positions destination (seconde colonne). La troisième colonne montre les positions des points pris dans l'espace origine après leur avoir appliqué notre transformation $T_{OVD}$ . Chaque ligne expose un cas. La seconde ligne reprend le cas de la première ligne après avoir appliqué une transformation $T_O$ aux positions d'origine alors que pour la troisième ligne une transformation $T_D$ est appliquée préalablement aux positions destination. Nous remarquons dans la troisième colonne que les positions résultantes sont invariantes à $T_O$ et covariantes avec $T_D$ . . . . .	98
3.5	Invariance et covariance des coordonnées $R\hat{V}B$ obtenues par application de la transformation $T_{IVC}$ . Voir le texte pour les explications.	100
3.6	Le descripteur IVC est constitué des coordonnées 3D des 12 coins rouges . Cf. le texte pour plus de détails. . . . .	103
3.7	Invariance et covariance des coordonnées $\hat{x}\hat{y}$ obtenues après application de la transformation $T_{CVI}$ . Cf. le texte pour plus de détails. .	110
3.8	Le descripteur CVI est constitué de $3 \times 16$ valeurs extraites des 16 transformations 2D représentées par les flèches rouges. Cf. le texte pour plus de détails. . . . .	112

4.1	Les pixels caractérisés par un niveau d'intensité de vert égal à 50 dans la première image sont présentés en blanc. La seconde ligne représente pour chaque image, l'histogramme vert de ces pixels représentés en blanc. Nous observons dans ces histogrammes que tous ces pixels n'ont pas tous le même niveau mais que leurs niveaux sont répartis autour d'une valeur. Ceci est expliqué pour une part par le bruit d'acquisition mais aussi et surtout par le phénomène de métamérisme. . . . .	120
4.2	Les histogrammes 2D rouge-vert des rectangles $r_1$ et $r_2$ . . . . .	123
4.3	Deux régions d'intérêts correspondantes détectées dans deux images acquises sous différentes conditions d'éclairage. . . . .	127
5.1	Les images 5.1(a), 5.1(h) et 5.1(o) représentent le même symbole mais ne possèdent aucune caractéristique locale de couleur ou de texture similaires. Les régions locales de référence 1 (5.1(b), 5.1(i), 5.1(p)), 2 (5.1(d), 5.1(k), 5.1(r)) et 3 (5.1(f), 5.1(m) et 5.1(t)) sont extraites à partir de positions correspondantes dans les 3 images. Les figures 5.1(c), 5.1(j), 5.1(q), 5.1(e), 5.1(l), 5.1(s), 5.1(g), 5.1(n) et 5.1(u) représentent les similarités (évaluées par notre méthode) entre les régions de référence et leur image respective. . . . .	132
5.2	Extraction du descripteur local d'auto-similarité. Cette figure a été empruntée à [95] . . . . .	134
5.3	Les descripteurs locaux extraits autour de 3 régions correspondantes. Cette figure a été empruntée à [95] . . . . .	136
5.4	Répartitions des auto-similarités globales de deux régions dans chaque image. Cette figure a été empruntée à [24] . . . . .	137

5.5	Les images 5.5(c) et 5.5(d) représentent les auto-similarités globales entre la région 5.5(b) extraites de l'image 5.5(a) et toutes les régions de cette même image. Les auto-similarités de l'image 5.5(c) sont déduites des différences pixel à pixel entre les régions alors que celles de l'image 5.5(d) sont obtenues grâce à notre approche. . . . .	140
6.1	Deux objets de la base ALOI sous différentes conditions. . . . .	153
6.2	Les conséquences des variations artificielles sur les couleurs de deux images. . . . .	154
6.3	Quelques images de la base Flowers . . . . .	155
6.4	Quelques images de la base Birds . . . . .	156
6.5	Quelques images de la base Soccers . . . . .	157
6.6	Quelques images de la base VOC 2010 . . . . .	158
6.7	La taille de CVI sur la base ALOI . . . . .	174
6.8	La taille d'IVC sur la base ALOI . . . . .	175
6.9	La normalisation d'IVC sur la base ALOI . . . . .	176
6.10	La normalisation d'IVC sur la base ALOI (perturbations artificielles). Les conditions $C_1$ à $C_7$ sont décrites dans le tableau 6.1. . . . .	177
6.11	IVC 3D et 2D avec et sans contrainte . . . . .	178
6.12	Performances des descripteurs locaux sur la base ALOI . . . . .	179
6.13	Performances des descripteurs locaux sur la base ALOI (perturbations artificielles). Les conditions $C_1$ à $C_7$ sont décrites dans le tableau 6.1. . . . .	180
6.14	Précision moyenne obtenue par les descripteurs pour la classification sur 4 bases . . . . .	184
6.15	Précisions obtenues par les descripteurs d'auto-similarités sur la base VOC 2007. MAP est le pourcentage de la précision moyenne. . . . .	185
A.1	Des images entrées à sac de mots visuels. . . . .	215



B.1	Descripteur SIFT correspondant à une région d'intérêt, la fenêtre d'étude constitué de 16 zones associée à ce descripteur est orientée selon la direction dominante selon laquelle le descripteur SIFT est calculé. . . . .	218
-----	---	-----

---

## LISTE DES TABLEAUX

2.1	Caractéristiques couleur invariantes présentées dans le paragraphe précédent . . . . .	64
2.2	Hypothèses de base utilisées par les caractéristiques couleur . . .	65
2.3	Hypothèses sur lesquelles reposent chaque caractéristique couleur invariante . . . . .	66
2.4	Phénomènes photométriques et radiométriques auxquels sont in-variantes chaque caractéristique . . . . .	67
6.1	Valeurs des paramètres de la transformation affine de l'équation (6.1) pour les conditions $C_v$ , $v = 1, \dots, 7$ des 4 tests. . . . .	151
6.2	Dimensions des descripteurs testés . . . . .	160
6.3	Evolution des courbes de PAC en fonction de l'invariance et du pouvoir discriminant du descripteur testé. . . . .	163



# **Annexes**

---

---

# ANNEXE A

---

## SAC DE MOTS

Selon l'approche sac de mots [66], chaque image est représentée par un sac de mots. Pour avoir un sac de mots, on doit tout d'abord créer un dictionnaire de mots visuels. Ce dictionnaire est obtenu par un (des) détecteur(s), un (des) descripteur(s) représenté sous la forme d'un vecteur et un algorithme qui regroupe le(s) vecteur(s) qui ont été définis par un (des) descripteur(s).

**Mot visuel** c'est une description représentative d'un ensemble de descripteurs invariants. Étant donné qu'un descripteur invariant est représenté par un vecteur, la présentation d'un mot visuel est aussi vectorielle. D'une manière générale, le vecteur moyen (description moyenne) d'un ensemble de descripteurs invariants est considéré comme le mot visuel le plus représentatif de cet ensemble.

**Sac de mots** c'est un ensemble de mots visuels qui apparaissent dans une image, on s'intéresse dans ce contexte seulement à quelques mots existants dans une image ; l'ordre de mots dans l'image n'a pas d'importance.

**Dictionnaire de mots visuels** c'est un ensemble de tous les mots visuels pos-

sibles. Tous les mots associés à un sac de mots doivent exister dans le dictionnaire.

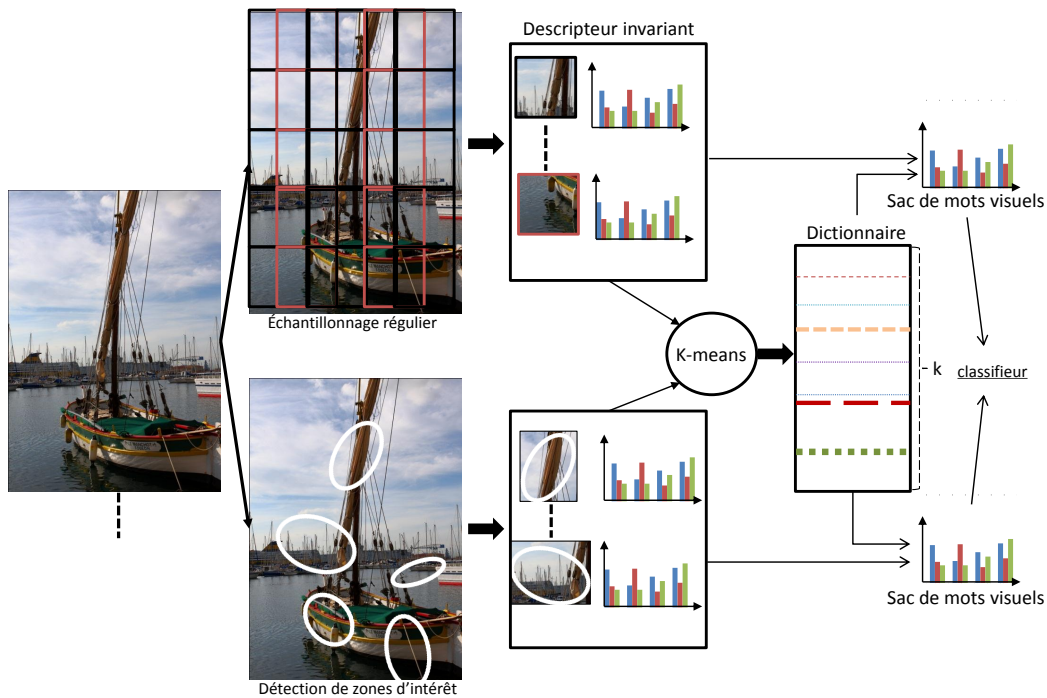


FIGURE A.1 – Des images entrées à sac de mots visuels.

L'illustration de la construction de sac de mots est présentée à la figure A.1. Cette figure illustre les différentes étapes de calculs des mots visuels et sacs de mots. Cette caractéristique d'une image opératoire est souvent structurée par les étapes suivantes :

1. Détecter sur les images entrées des régions régulières ou d'intérêts par Harris laplace, MSER ou SIFT.
2. Calculer des descripteurs invariants sur les régions définies et coder ces descripteurs sous forme de vecteur.

3. Construction du dictionnaire de mots visuels par l'algorithme des K-means pour former un dictionnaire de K mots visuels.
4. Indexer chaque image à partir des descripteurs calculés pour cette image en fonction du dictionnaire de mots visuels.
5. Classification d'images en fonction des sacs de mots visuels qui les caractérisent.

---

---

## ANNEXE B

---

# SIFT : SCALE-INVARIANT FEATURE TRANSFORM

L'approche SIFT (Scale-invariant feature transform) est lancée sur un algorithme complet qui a de détection à l'appariement de points d'intérêts, cette approche a été introduite par Lowe [68]. Dans cette annexe, nous allons uniquement nous focaliser sur la description d'une région locale par l'approche SIFT. Aucune introduction du détecteur SIFT n'est présentée ici. Supposons que suite à un premier prétraitement nous ayons déjà détecté une région locale d'intérêt par la méthode SIFT, MSER ou Harris-laplace. Le calcul du descripteur SIFT repose sur plusieurs étapes : 1, A chaque région locale d'intérêt (cf. Figure B.1) est associée une ou plusieurs orientations déterminées localement à partir de la direction des gradients calculées dans des zones images déterminées. Dans la mesure où les descripteurs SIFT sont calculés relativement à ces orientations, cette étape est essentielle pour garantir l'invariance de ces descripteurs à la rotation. Il est en effet fondamental que ce descripteur puisse garder une valeur constante, quelle que soit l'orientation de l'image. Il est important de souligner que cette première



étape peut être utilisée également pour définir d'autres descripteurs invariants à la rotation. On peut donc considérer comme une étape préalable dans la description d'un descripteur invariant à la rotation.

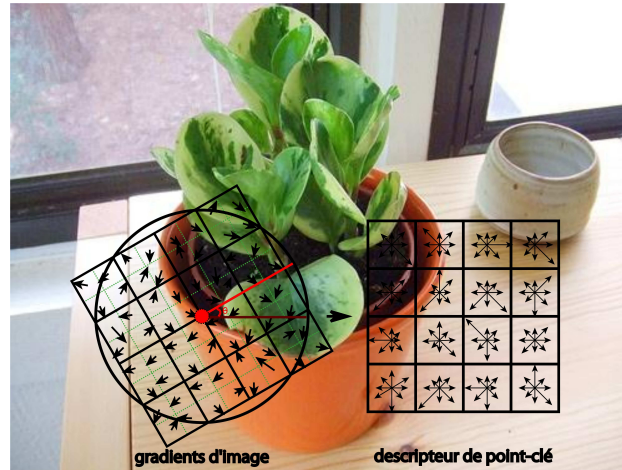


FIGURE B.1 – Descripteur SIFT correspondant à une région d'intérêt, la fenêtre d'étude constitué de 16 zones associée à ce descripteur est orientée selon la direction dominante selon laquelle le descripteur SIFT est calculé.

Cette seconde étape est réalisée sur une version lissée de l'image en fonction d'un facteur d'échelle prédéfini. Chaque région locale d'intérêt est subdivisée en 4 zones, sur chaque zone est calculé un histogramme des orientations quantifié sur 8 intervalles de valeurs. En chaque point de ces zones, l'orientation ( $\theta$ ) et la magnitude ( $m$ ) du gradient de luminance ( $L$ ) sont calculés comme indiqué dans l'équation B.1. Afin de donner plus de poids aux zones centrales de la zone d'intérêt (i.e les zones les plus proches du point rouge dans la Figure B.1), un masque gaussien est appliqué à la magnitude. Ensuite, les 16 histogrammes codés sur 8 intervalles chacun sont concaténés puis normalisés. Dans le but de diminuer la sensibilité du descripteur aux changements de luminosité, les valeurs supérieures à 0.2 sont remplacées par la valeur 0.2 et l'histogramme est de nouveau normalisé, ce qui nous donne au final le descripteur SIFT de la région locale, le quel

descripteur est de dimension 128 (i.e  $16 \times 8$ ).

$$\begin{aligned} m(x, y) &= \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \\ \theta &= \tan^{-1} \left( \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right) \end{aligned} \quad (\text{B.1})$$

---

---

## ANNEXE C

---

### LA LISTE DES PUBLICATIONS EN RAPPORT AVEC LA THÈSE

- X.Song, D.Muselet and A.Trémeau. Local Color Descriptor for Object Recognition Across Illumination Changes, *Proc. Of Advanced Concepts for Intelligent Vision systems*, vol 8, Bordeaux, France, 2009
- X.Song, D.Muselet and A.Trémeau. Compact local color descriptor based on rank correlations, *Color and Reflectance in Imaging and Computer Vision Workshop 2009*, vol 7, Kyoto, Japan, 2009,
- M.Seifi, X.Song, D.Muselet and A.Trémeaux. Color texture classification across illumination changes, *Proc. Of International Conference on Computer Graphics*, vol 6, Joensuu, Finland, 2010
- D.Muselet, X.Song, and L.Macaire, Exploitation de la couleur pour la reconnaissance d'objets, *Proc. Of Ecole d'Hiver*. vol11, toulon, France, 2009
- D.Muselet, X.Song, Chapitre "Les caractéristiques couleur invariantes pour

## ANNEXE C. LA LISTE DES PUBLICATIONS EN RAPPORT AVEC LA THÈSE

---

la reconnaissance d'objets" *Hermes Science Publishing Ltd*, France, 2012

- Un article portant sur le chapitre 3 sera soumis à la revue IEEE Transactions on Pattern Analysis and Machine Intelligence avant la fin du mois d'octobre.
- Un article portant sur le chapitre 5 sera soumis dans une conférence internationale avant la fin de l'année.

ANNEXE C. LA LISTE DES PUBLICATIONS EN RAPPORT AVEC LA  
THÈSE

---